PRINCIPLES OF NUTRITIONAL ASSESSMENT 3E

Rosalind S. Gibson et al.
University of Otago

LibreTexts™

Principles of Nutritional Assessment: 3rd Edition, April 2024

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden borne from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and were supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 09/13/2026

TABLE OF CONTENTS

Licensing

1: Introduction

1.1: New developments in nutritional assessment

1.2: Nutritional assessment systems

1.3: Nutritional assessment methods

1.4: Nutritional assessment indices and indicators

1.5: The design of nutritional assessment systems

1.6: Important characteristics of assessment measures

1.7: Evaluation of nutritional assessment indices

2: Food available for consumption at National and Household Levels

2.1: Introduction

2.2: Food Balance Sheets

2.3: Household Surveys

2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)

2.5: National and household surveys in the global assessment of food security and diet quality by age and sex

3: Food consumption of individuals

3.1: Methods for measuring food consumption of individuals

3.2: Technical improvements in food consumption measurements

3.3: Selecting an appropriate method

4: Measurement Error in Dietary assessment (Chapter 5)

4.1: Measurement error in dietary intake data

4.2: Sources of measurement error

4.3: Implications of measurement error for estimated dietary intakes

4.4: Minimizing measurement error through data collection procedures

4.5: Summary

5: Nutrient reference values (Chapter 8a)

5.1: Development of Nutrient Reference Values (8a.1)

5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)

5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)

5.4: U.K. Dietary Reference Values (8a.4)

5.5: United States and Canada- Dietary Reference Intakes (8a.5)

5.6: European Dietary Reference Values (8a.6)

5.7: WHO/FAO Nutrient Reference Values (8a.7)

5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)

6: Evaluation of intakes and diets (Chapter 8b)

6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

6.4: Using the NRVs for planning nutritious diets (8b.4)

7: Dietary guidelines and quality (Chapter 8c)

7.1: Food-based dietary guidelines (8c.1)

7.2: Assessing diet quality (8c.2)

7.3: Examples of diet quality indices (8c.3)

7.4: Alternate Healthy Eating Index (8c.4)

7.5: Mediterranean Diet Score or Index (8c.5)

7.6: Dietary Approaches to Stop Hypertension scores

7.7: Healthy Diet Indicator (8c.7)

7.8: Dietary Inflammatory Indices (8c.8)

7.9: Diet quality based on level of food processing (8c.9)

7.10: Lower-burden diet quality indices (8c.10)

7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)

7.12: Summary of diet quality indices (8c.12)

8: Introduction to Anthropometry (Chapter 9)

8.1: Measurements, indices, and indicators

8.2: Advantages and limitations of anthropometry

8.3: Errors in anthropometry

8.4: Interpretation and evaluation of anthropometric data

9: Body size (Chapter 10)

9.1: Anthropometric assessment of body size (10.0)

9.2: Measurements of body size (10.1)

9.3: Growth indices, indicators, and recommended growth reference data (10.2)

9.4: Body mass index in adults (10.3)

9.5: BMI in children and adolescents (10.4)

10: Body composition (Chapter 11)

10.1: Anthropometric assessment of body composition (11.0)

10.2: Assessment of body fat (11.1)

10.3: Assessment of fat-free mass (11.2)

11: Evaluation of Anthropometric Data (Chapter 13)

11.1: Introduction (13.0)

11.2: Anthropometric indices and their modes of expression (13.1)

11.3: Use of anthropometric indices in clinical settings (13.2)

11.4: Screening systems to identify individuals at risk in a population (13.3)

11.5: Applications of anthropometry in public health (13.4)

12: Body Composition- Laboratory Methods (Chapter 14)

12.1: An introduction to techniques used to measure body composition (14.0)

12.2: Chemical analysis of cadavers (14.1)

12.3: Total body potassium (TBK) (14.2)

12.4: Total body water from isotope dilution (14.3)

12.5: Multiple dilution methods (14.4)

12.6: In vivo activation analysis (14.5)

12.7: Densitometry (14.6)

12.8: Total body electrical conductivity (14.7)

12.9: Bioelectrical impedance (14.8)

12.10: Computerized tomography (14.9)

12.11: Magnetic resonance imaging (14.10)

12.12: Dual energy X-ray absorptiometry (14.11)

12.13: Ultrasound (14.12)

12.14: Summary - Method comparisons (14.13)

13: Biomarkers (Chapter 15)

13.1: Biomarkers to assess nutritional status (15.1)

13.2: Biomarkers of exposure (15.2)

13.3: Biomarkers of status (15.3)

13.4: Biomarkers of function (15.4)

13.5: Factors affecting choice of nutritional biomarkers (15.5)

13.6: Evaluation of the selected nutritional biomarker (15.6)

13.7: Application of new technologies (15.7)

14: Iron (Chapter 17)

14.1: Introduction and functions of iron (17.1)

14.2: Hemoglobin (17.2)

14.3: Hematocrit or packed cell volume

14.4: Red-cell indices

14.5: Red-cell distribution width (17.5)

14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)

14.7: Serum ferritin (17.7)

14.8: Erythrocyte protoporphyrin

14.9: Serum soluble transferrin receptor (17.9)

14.10: Multiple biomarkers (17.10)

14.11: Emerging iron status indicators (17.11)

15: Vitamin A (Chapter 18a)

15.1: Introduction (18a.1)

15.2: Indices of vitamin A status (18a.2)

16: Vitamin D (Chapter 18b)

16.1: Functions of vitamin D (18b.1)

16.2: Metabolism of vitamin D (18b.2)

16.3: Vitamin D deficiency in humans (18b.3)

16.4: Food sources and dietary intakes (18b.4)

16.5: Nutrient reference values for vitamin D (18b.5)

16.6: Biomarkers of vitamin D status (18b.6)

16.7: Interpretive criteria (18b.8)

16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)

16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)

16.10: Other vitamin D metabolites (18b.11)

16.11: Serum parathyroid hormone (18b.13)

16.12: Calcium and phosphorus in serum and urine (18b.14)

17: Vitamin C (Chapater 19)

17.1: Vitamin C (19.1)

17.2: Functions of Vitamin C (19.2)

17.3: Absorption and metabolism of vitamin C (19.3)

17.4: Deficiency of vitamin C in humans (19.4)

17.5: Food sources and dietary intakes of vitamin C (19.5)

17.6: Nutrient reference values for vitamin C (19.6)

17.7: Indices of vitamin C status (19.7)

17.8: Plasma ascorbic acid (19.8)

17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)

17.10: Ascorbic acid in erythrocytes and whole blood (19.10)

17.11: Urinary excretion of ascorbic acid and metabolites (19.11)

17.12: Ascorbic acid in other cells and fluids (19.12)

17.13: Body pool size (19.13)

17.14: Functional tests of vitamin C status (19.14)

18: Thiamine (Chapter 20a)

18.1: Thiamine (20a.1)

18.2: Biochemical indices of thiamine status (20a.2)

18.3: Conclusions (20a.3)

19: Riboflavin (Chapter 20b)

19.1: Riboflavin (20b.1)

19.2: Biomarkers of riboflavin status (20b.2)

20: Niacin (Chapter 20c)

20.1: Niacin - Introduction (20c.1)

20.2: Indices of niacin status (20c.2)

21: Folate (Chapter 22a)

21.1: Folate (22a.1)

21.2: Biomarkers of folate status (Chapter 22a.2)

21.3: Conclusions (22a.3)

22: Vitamin B12 (Chapter 22b)

22.1: Introduction (22b.1)

22.2: Biomarkers of vitamin B12 status (22b.2)

23: Calcium (Chapter 23a)

23.1: Calcium (23a.1)

23.2: Functions of calcium (23a.2)

23.3: Absorption and metabolism of calcium (23a.3)

23.4: Dietary sources and intakes of calcium (23a.4)

23.5: Nutrient reference values for calcium (23a.5)

23.6: Calcium deficiency (23a.6)

23.7: Effects of high intakes of calcium (23a.7)

23.8: Biochemical indices of calcium status (23a.8)

23.9: Measurement of bone mineral content and bone mineral density (23a.9)

24: Phosphorus (Chapter 23b)

24.1: Phosphorus (23b.1)

24.2: Biological forms of phosphate and their measurement (23b.2)

24.3: Interpretive criterion- serum phosphate (23b.3)

24.4: Phosphate balance (23b.4)

24.5: Hormonal regulation of serum phosphate (23b.5)

24.6: Nutrient reference values (23b.6)

24.7: Dietary sources

24.8: Abnormalities in serum phosphate (23b.8)

24.9: Inherited and tumor-induced disorders of phosphate metabolism

25: Phosphorus (Chapter 23b)

25.1: Introduction (23c.1)

25.2: Functions of magnesium (23c.2)

25.3: Absorption and metabolism (23c.3)

25.4: Magnesium deficiency in humans

25.5: Food sources and dietary Intakes

25.6: Effects of high magnesium intakes (23c.6)

25.7: Occurrence of dietary deficiency (23c.7)

25.8: Indices of magnesium status (23c.8)

26: Magnesium (Chapter 23c)

26.1: Introduction (23c.1)

26.2: Functions of magnesium (23c.2)

26.3: Absorption and metabolism (23c.3)
26.4: Magnesium deficiency in humans (23c.4)
26.5: Food sources and dietary Intakes (23c.5)
26.6: Effects of high magnesium intakes (23c.6)
26.7: Occurrence of dietary deficiency (23c.7)
26.8: Indices of magnesium status (23c.8)

27: Copper (Chapter 24b)

27.1: Copper (24b.0)
27.2: Serum copper (24b.1)
27.3: Serum ceruloplasmin (24b.2)
27.4: Erythrocyte superoxide dismutase (24b.3)
27.5: Cytochrome c oxidase (24b.4)
27.6: Skin lysyl oxidase (24b.5)
27.7: Diamine oxidase (24b.6)
27.8: Copper chaperones (24b.7)
27.9: Other indices of copper status (24b.8)

28: Zinc (Chapter 24c)

28.1: Zinc (24c.1)
28.2: Dietary Zinc Intake (24c.2)
28.3: Serum zinc (24c.3)
28.4: Linear growth (24c.4)
28.5: Urine zinc (24c.5)
28.6: Hair zinc concentrations (24c.6)
28.7: Nail zinc (24c.7)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
28.11: Oxidative stress and DNA integrity (24c.11)
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
28.13: Taste acuity (24c.13)
28.14: Zinc-dependent enzymes (24c.14)
28.15: Erythrocyte zinc (24c.15)
28.16: Leukocyte zinc (24c.16)
28.17: Salivary zinc (24c.17)
28.18: Multiple biomarkers (24c.18)

29: Selenium (Chapter 25b)

29.1: Overview (25b.1)
29.2: Selenium Chemistry (25b.2)
29.3: Health Effects (25b.3)
29.4: Selenium Requirements (25b.4)

29.5: Food Sources (25b.5)
29.6: Enteric Absorption (25b.6)
29.7: Metabolism (25b.7)
29.8: Biochemical Functions (25b.8)
29.9: Assessing Selenium Status (25b.9)
29.10: Questions Remain (25b.10)

Index

Glossary

Detailed Licensing

This page titled Principles of Nutritional Assessment 3e (Gibson et al.) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson et al..

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

This page titled Licensing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1: INTRODUCTION

Nutritional assessment procedures were first used in surveys designed to describe the nutritional status of populations on a national basis. The assessment methods used were initially described following a conference held in 1932 by the Health Organization of the League of Nations.

In 1955, the Interdepartmental Committee on Nutrition for National Defense (ICNND) was organized to assist low-income countries in assessing the nutritional status of their populations and to identify problems of malnutrition and the ways in which they could be solved. The ICNND teams conducted medical nutrition surveys in 24 countries. A comprehensive manual was then produced and later, updated guidance issued (ICNND, 1984) with the intention of standardizing both the assessment methods used for the collection of nutrition survey data and the interpretation of the results.

On the recommendation of a World Health Organization (WHO) Expert Committee on Medical assessment of Nutritional Status, a second publication was prepared by Jelliffe (1966) in consultation with 25 specialists from various countries. This monograph was directed specifically at the assessment of the nutritional status of vulnerable groups in low-income countries of the world.

Many of the methods described in this monograph are still used by the the U.S. Demographic and Health Surveys (DHS) Program to collect representative data on population, health, HIV, and nutrition, and about 30 indicators supporting the Sustainable Development Goals. The data are used to identify public health nutrition problems so that effective intervention programs can be designed. The U.S. DHS program has conducted more than 400 surveys in over 90 low- and middle-income countries since 1984.

Many higher income countries collect national data on the nutritional status of the population, some (e.g., the U.S. and the U.K.) collecting data on an ongoing basis using nutrition surveillance systems. In the past, these systems have often targeted high-risk populations, especially low-income mothers, children under five, and pregnant women. Now, with the growing awareness of the role of nutrition as a risk factor for chronic diseases, surveillance systems often encompass all age groups.

1.1: New developments in nutritional assessment

1.2: Nutritional assessment systems

1.3: Nutritional assessment methods

1.4: Nutritional assessment indices and indicators

1.5: The design of nutritional assessment systems

1.6: Important characteristics of assessment measures

1.7: Evaluation of nutritional assessment indices

This page titled 1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1.1: New developments in nutritional assessment

Today, nutritional assessment emphasizes new simple noninvasive approaches, particularly valuable in low income countries, that can measure the risk of nutrient deficits and excesses, and monitor and evaluate the effects of nutrition interventions. These new approaches to assessment include the measurement of nutrients and biomarkers in dried blood spots prepared from a finger-prick blood sample, avoiding the necessity for venous blood collection and refrigerated storage (Mei et al., 2001). In addition, for some nutrients, on-site analysis is now possible, enabling researchers and respondents to obtain results immediately.

Many of these new approaches can also be applied to biomarkers monitoring the risk of chronic diseases. These include biomarkers of antioxidant protection, soft-tissue oxidation, and free-radical formation, all of which have numerous clinical applications.

Increasingly, "all-in-one" instrumental platforms for multiple micronutrient tests on a single sample aliquot are being developed, some of which have been adapted for dried blood spot matrices (Brindle et al., 2019). These assessment instruments are designed so they are of low complexity and can be operated by laboratory technicians with minimal training, making them especially useful in low- and middle-income countries (Esmaeili et al., 2019).

The public availability of e- and m-Health communication technologies has increased dramatically in recent years. e-Health is defined as:

> "the use of emerging information and communications technology, especially the internet, to improve or enable health and health care"

whereas m-Health interventions are

> "those designed for delivery through mobile phones" (Olson, 2016).

Interventions using these communication technologies to assess, monitor and improve nutrition-related behaviors and body weight, appear to be efficacious across cognitive outcomes, and some behavioral and emotional outcomes, although changing dietary behaviors is a more challenging outcome. There is an urgent need for a rigorous scientific evaluation of e- and m-health intervention technologies. To date their public health impact remains uncertain.

Nutritional assessment is also an essential component of the nutritional care of the hospitalized patient. The important relationship between nutritional status and health, and particularly the critical role of nutrition in recovery from acute illness or injury, is well documented. Although it is many years since the prevalence of malnutrition among hospitalized patients was first reported (Bistrian et al., 1974, 1976), such malnutrition still persists (Barker et al., 2011).

In the early 1990s, evidence-based medicine started as a movement in an effort to optimize clinical care. Originally, evidence-based-medicine focused on critical appraisal, followed by the development of methods and techniques for generating systematic reviews and clinical practice guidelines (Djulbegovic and Guyatt, 2017). For more details see Section 1.2.6.

Point of care technology (POCT) is also a rapidly expanding health care approach that can be used in diverse settings, particularly those with limited health services or laboratory infrastructure, as the tests do not require specialized equipment and are simple to use. The tests are also quick, enabling prompt clinical decisions to be made to improve the patient's health at or near the site of patient care. The development and evaluation of POC devices for the diagnosis of malaria, tuberculosis, HIV, and other infectious diseases is on-going and holds promise for low-resource settings(Heidt et al., 2020; Mitra and Sharma, 2021). Guidelines by WHO (2019) for the development of POC devices globally are available, but challenges with regulatory approval, quality assurance programs, and product service and support remain (Drain et al., 2014).

Personalized nutrition is also a rapidly expanding approach that tailors dietary recommendation to the specific biological requirements of an individual on the basis of their health status and performance goals. See Setion 1.1.5 for more details. The approach has become possible with the increasing advances in "-omic sciences" (e.g., nutrigenomics, proteomics and metabolomics). See Chapter 15 and van Ommen et al. (2017) for more details.

Health-care administrators and the community in general, continue to demand demonstrable benefits from the investment of

public funds in nutrition intervention programs. This requires improved techniques in nutritional assessment and the monitoring and evaluation of nutrition interventions. In addition, implementation research is now being recognized as critical for maximizing the benefits of evidence-based interventions. Implementation research in nutrition aims to build evidence-based knowledge and sound theory to design and implement programs that will deliver nutrition programs effectively. However, to overcome the unique challenges faced during the implementation of nutrition and health interventions, strengthening the capacity of practitioners alongside that of health researchers is essential. Dako-Gyke et al. (2020) have developed an implementation research course curriculum that targets both practitioners and researchers simultaneously, and which is focused on low- and middle-income countries.

The aim of this 3rd edition of "Principles of Nutritional assessment" is to provide guidance on some of these new, improved techniques, as well as a comprehensive and critical appraisal of many of the classic, well-established methods in nutritional assessment.

This page titled 1.1: New developments in nutritional assessment is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1.2: Nutritional assessment systems

Nutritional assessment systems involve the interpretation of information from dietary and nutritional biomarkers, and anthropometric and clinical studies. The information is used to determine the nutritional status of individuals or population groups as influenced by the intake and utilization of dietary substances and nutrients required to support growth, repair, and maintenance of the body as a whole or in any of its parts (Raiten and Combs, 2015).

Nutritional assessment systems can take one of four forms: surveys, surveillance, screening, or interventions. These are described briefly below.

1.2.1 Nutrition surveys

The nutritional status of a selected population group is often assessed by means of a cross-sectional survey. The survey may either establish baseline nutritional data or ascertain the overall nutritional status of the population. Cross-sectional nutrition surveys can be used to examine associations, and to identify and describe population subgroups "at risk" for chronic malnutrition. Causal relationships cannot be established from cross-sectional surveys because whether the exposure precedes or follows the effect is unknown. They are also unlikely to identify acute malnutrition because all the measurements are taken on a single occasion or within a short time period with no follow-up. Nevertheless, information on prevalence, defined as the proportion who have a condition or disease at one time point, can be obtained from cross-sectional surveys for use by health planners. Cross-sectional surveys are also a necessary and frequent first step in subsequent investigations into the causes of malnutrition or disease.

National cross-sectional nutrition surveys generate valuable information on the prevalence of existing health and nutritional problems in a country that can be used both to allocate resources to those population subgroups in need, and to formulate policies to improve the overall nutrition of the population. They are also sometimes used to evaluate nutrition interventions by collecting baseline data before, and at the end of a nutrition intervention program, even though such a design is weak as the change may be attributable to some other factor (Section 1.2.4).

Several large-scale national nutrition surveys have been conducted in industrialized countries during the last decade. They include surveys in the United States, the United Kingdom, Ireland, New Zealand, and Australia. More than 400 Demographic and Health Surveys (DHS) in over 90 low- and middle-income countries have also been completed. See U.S. DHS program.

1.2.2 Nutrition surveillance

The characteristic feature of surveillance is the continuous monitoring of the nutritional status of selected population groups. Surveillance studies therefore differ from nutrition surveys because the data are collected, analyzed, and utilized over an extended period of time. Sometimes, the surveillance only involves specific at-risk subgroups, identified in earlier nutrition surveys.

The information collected from nutrition surveillance programs can be used to achieve the objectives shown in Box 1.1.

Box 1.1 Objectives of nutrition surveillance

- Aid long-term planning in health and development;
- Provide input for program management and evaluation;
- Give timely warning of the need for intervention to prevent critical deteriorations in food consumption.

To achieve these objectives, the nutrition information collected must be:

- population-based;
- decision and action orientated;
- sensitive and accurate;

- relevant and timely;
- readily accessible;
- communicated effectively.

Modified from Jerome and Ricci (1997).

Surveillance studies, unlike cross-sectional nutrition surveys, can also identify the possible causes of both chronic and acute malnutrition and, hence, can be used to formulate and initiate intervention measures at either the population or the subpopulation level.

In the United States, a comprehensive program of national nutrition surveillance, known as the National Health and Nutrition Examination Survey(NHANES), has been conducted since 1959. Data on anthropometry, demographic and socio-economic status, dietary and health-related measures are collected. In 2008, the United Kingdom began the National Diet and Nutrition Survey Rolling Program. This is a continuous program of field work designed to assess the diet, nutrient intake, and nutritional status of the general population aged 1.5y and over living in private households in the U.K. (Whitton et al., 2011). WHO has provided some countries with surveillance systems so that they can monitor changes in the global targets to reduce the high burden of disease associated with malnutrition.

Note that the term "nutrition monitoring," rather than nutrition surveillance, is often used when the participants selected are high-risk individuals (e.g., food-insecure households, pregnant women). For example, because household food insecurity is of increasing public health concern, even in high-income countries such as the U.S. and Canada, food insecurity is regularly monitored in these countries using the Household Food Security Survey Module (HFSSM). Also see: Loopstra (2018).

1.2.3 Nutrition screening

The identification of malnourished individuals requiring intervention can be accomplished by nutrition screening. This involves a comparison of measurements on individuals with predetermined risk levels or "cutoff" points using measurements that are accurate, simple and cheap (Section 1.6.3), and which can be applied rapidly on a large scale. Nutrition screening can be carried out on the whole population, targeted to a specific subpopulation considered to be at risk, or on selected individuals. The programs are usually less comprehensive than surveys or surveillance studies.

Numerous nutrition screening tools are available for the early identification and treatment of malnutrition in hospital patients and nursing homes, of which Subjective Global assessment (SGA) and the Malnutrition Universal Screening Tool (MUST) are widely used; see Barker et al. (2011) and Chapter 27 for more details.

In low-income countries, mid-upper-arm circumference (MUAC) with a fixed cutoff of 115mm is often used as screening tool to diagnose severe acute malnutrition (SAM) in children aged 6–60mos (WHO/UNICEF, 2009). In some settings, mothers have been supplied with MUAC tapes either labeled with a specific cutoff of < 115mm, or color-coded in red (MUAC < 115mm), yellow (MUAC = 115–124mm), and green (MUAC > 125mm) in an effort to detect malnutrition early, before the onset of complications, and thus reduce the need for inpatient treatment (Blackwell et al., 2015; Isanaka et al., 2020).

In the United States, screening is used to identify individuals who might benefit from the Supplemental Nutrition Assistance Program (SNAP). The program is means tested with highly selective qualifying criteria. The SNAP (formerly food stamps) program provides money loaded onto a payment card which can be used to purchase eligible foods, to ensure that eligible households do not go without foods. In general, studies have reported that participation in SNAP is associated with a significant decline in food insecurity (Mabli and Ohls, 2015).

The U.S. also has a Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) that targets low-income pregnant and post-partum women, infants, and children < 5y. In 2009, the USDA updated the WIC food packages in an effort to balance nutrient adequacy with reducing the risk of obesity; details of the updates are available in NASEM (2006). Guthrie et al. (2020) compared associations between WIC participants and the nutrients and food packages consumed in 2008 and in 2016 using data from cross-sectional nationwide surveys of children aged < 4y. The findings indicated that more WIC infants who received the updated WIC food packages in 2016 had nutrient intakes (except iron) that met their estimated average requirements (EARs). Moreover, vegetables provided a larger contribution to their nutrient intakes, and intakes of low-fat milks had increased for toddlers aged 2y, likely contributing to their lower reported intakes of saturated fat.

1.2.4 Nutrition interventions

Nutrition interventions often target population subgroups identified as “at-risk” during nutrition surveys or by nutrition screening.

In 2013, the Lancet Maternal and Child Nutrition Series recommended a package of nutrition interventions that, if scaled to 90% coverage, could reduce stunting by 20% and reduce infant and child mortality by 15% (Bhutta et al., 2013). The nutrition interventions considered included lipid-based and micronutrient supplementation, food fortification, promotion of exclusive breast feeding, dietary approaches, complementary feeding, and nutrition education. More recently, nutrition interventions that address nutrition-sensitive agriculture are also being extensively investigated (Sharma et al., 2021)

Increasingly, health-care program administrators and funding agencies are requesting evidence that intervention programs are implemented as planned, reach their target group in a cost-effective manner, and are having the desired impact. Hence, monitoring and evaluation are becoming an essential component of all nutrition intervention programs. However, because the etiology of malnutrition is multi-factorial and requires a multi-sectorial response, the measurement and collection of the data from such multiple levels presents major challenges.

Several publications are available on the design, monitoring, and evaluation of nutrition interventions. The reader is advised to consult these sources for further details (Habicht et al., 1999; Rossi et al., 1999; Altman et al., 2001). Only a brief summary is given below.

Monitoring, discussed in detail by Levinson et al. (1999), oversees the implementation of an intervention, and can be used to assess service provision, utilization, coverage, and sometimes the cost of the program. Effective monitoring is essential to demonstrate that any observed result is probably from the intervention.

Emphasis on the importance of designing a program theory framework and associated program impact pathway (PIP) to understand and improve program delivery, utilization, and the potential of the program for nutritional impact has increased (Olney et al., 2013; Habicht and Pelto, 2019). The construction of a PIP helps conceptualize the program and its different components (i.e., inputs, processes, outputs, and outcomes to impacts). Only with this information can issues in program design, implementation, or utilization that may have the potential to limit the impact of the program, be identified, and, in turn strengthened, so the impact of the program can be optimized. Program impact pathway analysis generally includes both quantitative and qualitative methods (e.g., behavior-change communication) to ascertain the coverage of an intervention.

An example of the multiple levels of measurements and data that were collected to optimize the impact of a “Homestead Food Production” program conducted in Cambodia are itemized in Box 1.2. Three program impact pathways were hypothesized, each requiring the measurements of a set of input, process, and output indicators; for more details of the indicators measured, see Olney et al. (2013).

Box 1.2 Example of the three hypothesized program impact pathways

- Pathway 1: Increasing the availability of micronutrient-rich foods through increased household production of these foods (production-consumption pathway)
- Pathway 2: Income generation through the sale of products from the homestead food production program (production–income pathway)
- Pathway 3: Increased knowledge and adoption of optimal nutrition practices, including intake of micronutrient-rich foods (knowledge–adoption of optimal health- and nutrition-related practices pathway) and improve delivery, utilization, and potential for impact of a Homestead Food Production Program in Cambodia.

From Olney et al. (2013).

Program impact pathway analysis can also be used to ascertain the coverage of an intervention. Bottlenecks at each sequential step along the PIP can be identified along with the potential determinants of the bottlenecks (Habicht and Pelto, 2019). Coverage can be measured at the individual and at the population level; in the latter case, it is assessed as the proportion of beneficiaries who received the intervention at the specified quality level. Many of the nutrition interventions highlighted by Bhutta and colleagues (2013) in the Lancet Maternal and Child Nutrition Series have now been incorporated into national policies and programs in low- and middle-income countries. However, reliable data on their coverage are scarce, despite the importance of coverage to ensure sustained progress in reducing rates of malnutrition. In an effort to achieve this goal, Gillespie

et al. (2019)have proposed a set of indicators for tracking the coverage of high-impact nutrition-specific interventions which are delivered primarily through health systems, and recommend incorporation of these indicators into data collection mechanisms and relevant intervention delivery platforms. For more details, see Gillespie et al. (2019).

The evaluation of any nutrition intervention program requires the choice of an appropriate design to assess the performance or effect of the intervention. The choice of the design depends on the purpose of the evaluation and the level of precision required. For example, for large scale public health programs, based on the evaluation, decisions may be made to continue, expand, modify, strengthen, or discontinue the program; these aspects are discussed in detail by Habicht et al. (1999). The indicators used to address the evaluation objectives must also be carefully considered(Habicht & Pelletier, 1990; Habicht & Stoltzfus, 1997).

Designs used for nutrition interventions vary in their complexity; see Hulley et al. (2013) for more details. Three types of evaluation can be achieved from these designs: adequacy, plausibility and probability evaluation, each of which is addressed briefly below.

An adequacy evaluation is achieved when it has not been feasible to include a comparison or control group in the intervention design. Instead, a within-group design has been used. In these circumstances, the intervention is evaluated on the basis of whether the expected changes have occurred by comparing the outcome in the target group with either a previously defined goal, or with the change observed in the target group following the intervention program. An example might be distributing iron supplements to all the target group (e.g., all preschool children with iron deficiency anemia) and assessing whether the goal of < 10% prevalence of iron-deficiency anemia in the intervention area after two years, has been met. Obviously, when evaluating the outcome by assessing the adequacy of change over time, at least baseline and final measurements are needed. Note that because there is no control group in this design, any reported improvement in the group, even if it is statistically significant, cannot be causally linked to the intervention.

A plausibility evaluation can be conducted with several designs, including a nonrandomized between-group design, termed a quasi-experimental design in which the experimental group receives the intervention, but the control group does not. The design should preferably allow blinding (e.g., use an identical placebo). Because the participants are not randomized into the two groups, multivariate analysis is used to control for potential confounding factors and biases, although it may not be possible to fully remove these statistically. A between-group quasi-experimental design requires more resources and is therefore more expensive than the within-group design discussed earlier, and is used when decision makers require a greater degree of confidence that the observed changes are indeed due to the intervention program.

A probability evaluation, when properly executed, provides the highest level of evidence that the intervention caused the outcome, and is considered the gold standard method. The method requires the use of a randomized, controlled, double-blind experimental design, in which the participants are randomly assigned to either the intervention or the control group. Randomization is conducted to ensure that, within the limits of chance, the treatment and control groups will be comparable at the start of the study. In some randomized trials, the treatment groups are communities and not individuals, in which case they are known as “community” trials.

Figure 1.1 illustrates the importance of the participants being randomized to either the intervention or the control group when the control group outcomes have also improved as a result of nonprogram factors. Note that the intervention and control groups are similar at baseline in this figure as a result of randomization.

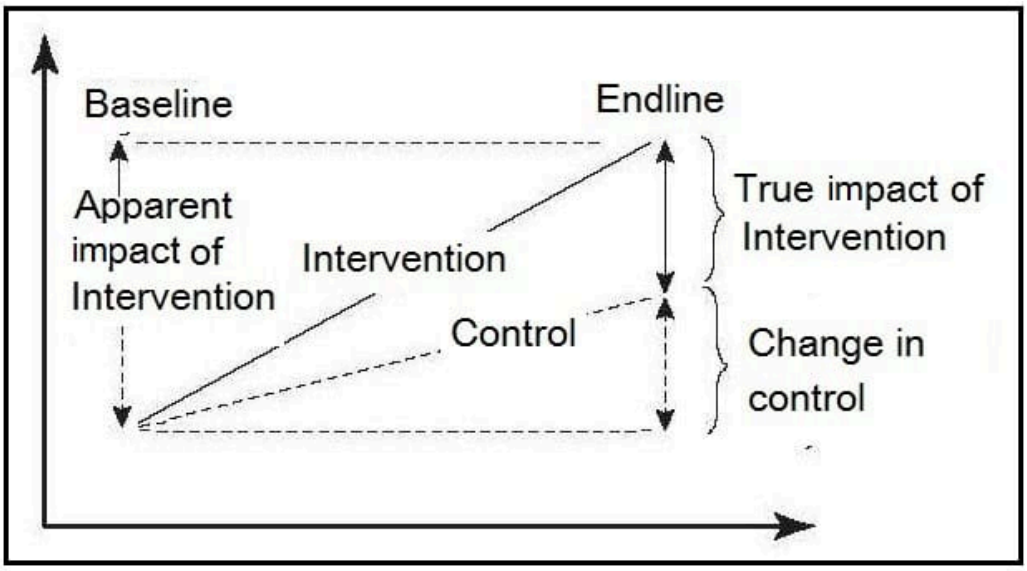


Figure 1.1. Example of the importance of a control group to distinguish true and apparent impact of an intervention in a scenario in which the outcomes in the control group have also improved as a result of nonprogram factors. Modified from Menon et al. (2013).

1.2.5 assessment systems in a clinical setting

The types of nutritional assessment systems used in the community have been adopted in clinical medicine to assess the nutritional status of hospitalized patients. This practice has arisen because of reports of the high prevalence of protein-energy malnutrition among surgical patients in North America and elsewhere (Corish and Kennedy, 2000; Barker et al., 2011). Today, nutritional assessment is often performed on patients with acute traumatic injury, on those undergoing surgery, on chronically ill medical patients, and on elderly patients. Initially, screening can be carried out to identify those patients requiring nutritional management. A more detailed and comprehensive baseline nutritional assessment of the individual may then follow. This assessment will clarify and expand the nutritional diagnosis, and establish the severity of the malnutrition. Finally, a nutrition intervention may be implemented, often incorporating nutritional monitoring and an evaluation system, to follow both the response of the patient to the nutritional therapy and its impact. Further details of protocols that have been developed to assess the nutritional status of hospital patients are given in Chapter 27.

Personalized nutrition is also a rapidly expanding approach that is being used in a clinical setting, as noted earlier. The approach tailors dietary recommendation to the specific biological requirements of an individual on the basis of their health status and performance goals. The latter are not restricted to the prevention and/or mitigation of chronic disease but often extend to strategies to achieve optimal health and well-being; some examples of these personal goals are depicted in Table 1.1. Personalized nutrition necessitates the use of a systems biology-based approach that considers the most relevant interacting biological mechanisms to formulate the best recommendations to meet the wellness goals of the individual.

Table 1.1. Examples of personal goals in relation to personal nutrition. Data from van Ommen et al. (2017).

Goal	Definition
Weight management	Maintaining (or attaining) an ideal body weight and/or body shaping that ties into heart, n
Metabolic health	Keeping metabolism healthy today and tomorrow
Cholesterol	Reducing and optimizing the balance between high-density lipoprotein and low-density li
Blood pressure	Reducing blood pressure in individuals who have elevated blood pressure
Heart health	Keeping the heart healthy today and tomorrow.
Muscle	Having muscle mass and muscle functional abilities. This is the physiological basis or unc
Endurance	Sustaining energy to meet the challenges of the day (e.g., energy to do that report at work
Strength	Feeling strong within yourself, avoiding muscle fatigue

Goal	Definition
Memory	Maintaining and attaining an optimal short-term and/or working memory
Attention	Maintaining and attaining optimal focused and sustained attention (i.e., being “in the mor

1.2.6 Approaches to evaluate the evidence from nutritional assessment studies

In an effort to optimize clinical care, evidence-based medicine (EBM) started as a movement in the early 1990s to enhance clinician's understanding, critical thinking, and use of the published research literature, while at the same time considering the patient’s values and preferences. It focused on the quality of evidence and risk of bias associated with the types of scientific studies used in nutritional assessment as shown in the EBM hierachy of evidence pyramid in Figure 1.2, with randomized controlled trials (RCTs) providing the strongest evidence and hence occupying the top tier.

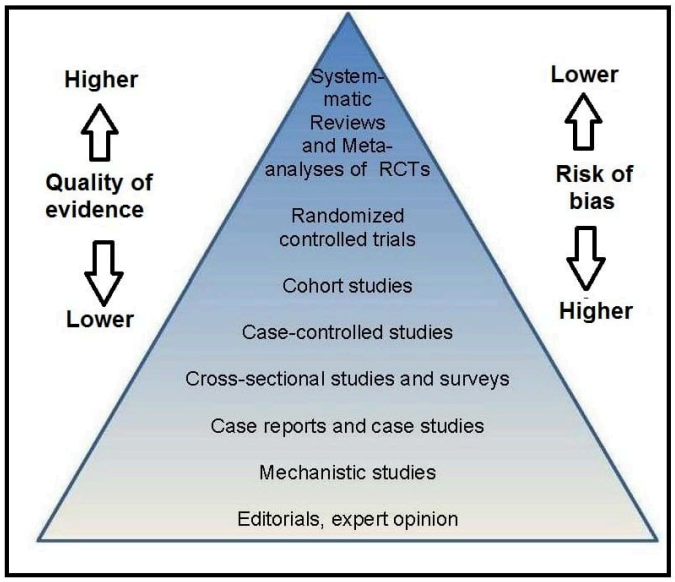


Figure 1.2. Traditional EBM hierarchy of evidence pyramid. The pyramidal shape qualitatively integrates the amount of evidence generally available from each type of study design and the strength of evidence expected from indicated designs. In each ascending level, the amount of available evidence generally declines. Study designs in ascending levels of the pyramid generally exhibit increased quality of evidence and reduced risk of bias. Confidence in causal relations increases at the upper levels. Meta-analyses and systematic reviews of observational studies and mechanistic studies are also possible. Redrawn from Yetley et al. (2017a).

Even within each level there are differences in the quality of evidence, depending on specific design features and conduct of the study. For example, seven bias domains are possible during the course of a study; these are shown in Figure 1.3. Of these bias domains, the four (i.e., numbers 4–7) that occur after the intervention has been assigned can operate in both randomized and nonrandomized study designs, whereas the other three (i.e., numbers 1–3)occur in observational studies and not in well-designed RCTs. Moreover, each bias specified in Figure 1.3 may contain several other different biases; see Hulley et al.(2013) and Yetley et al.(2017a) for more details.

Recognition of the importance of evaluating the evidence from individual studies has led to the development of three tools: Quality Assurance Instruments (QAIs), risk of bias tools, and an evidence-grading system. SIGN 50 is an example of a QAI that is widely used with versions available for cohort studies, case-control studies, and RCTs, and is based on a methodological checklist of items. In the future, QAIs will be available for nutrition studies based on RCTs, cohort, case-control, and cross-sectional studies with the aim of improving the consistency with which nutrition studies are assessed.

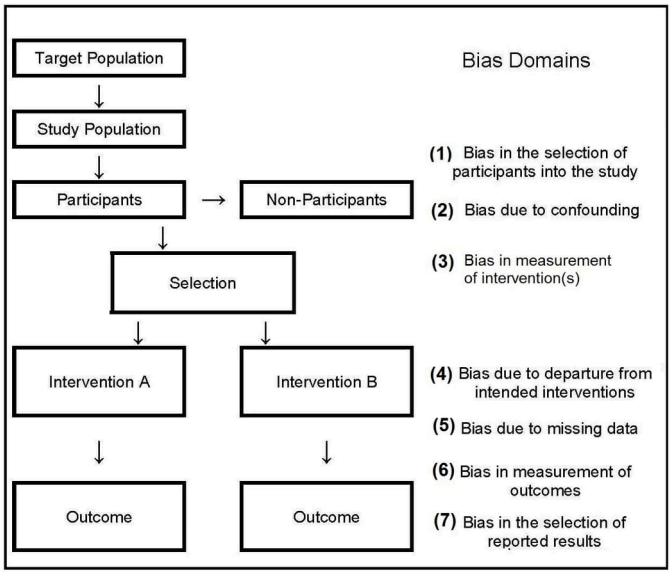


Figure 1.3. A flow chart of events that occur during a study with the seven different biases that can occur during the study. The biases are aligned with where in the study they occur. Redrawn from National Academies of Sciences, Engineering and Medicine (2018).

Risk of bias tools assess the degree of bias and are specific to study type. They focus on internal validity. Examples include the Cochrane Risk of Bias Tool used to evaluate RCTs (Cochrane Handbook), and the (ROBINS-I tool), best used to evaluate individual observational studies (Sterne et al., 2016). These tools assess six of the seven domains of bias listed in Figure 1.3, judging each as low, unclear, or high risk. A nutrition-specific risk of bias tool is in the planning stage.

Evidence-grading systems have also been developed for individual studies, of which the Grading of Recommendations, assessment, Development and Evaluation (GRADE) approach is widely used (Guyatt et al., 2011), Figure 1.4.

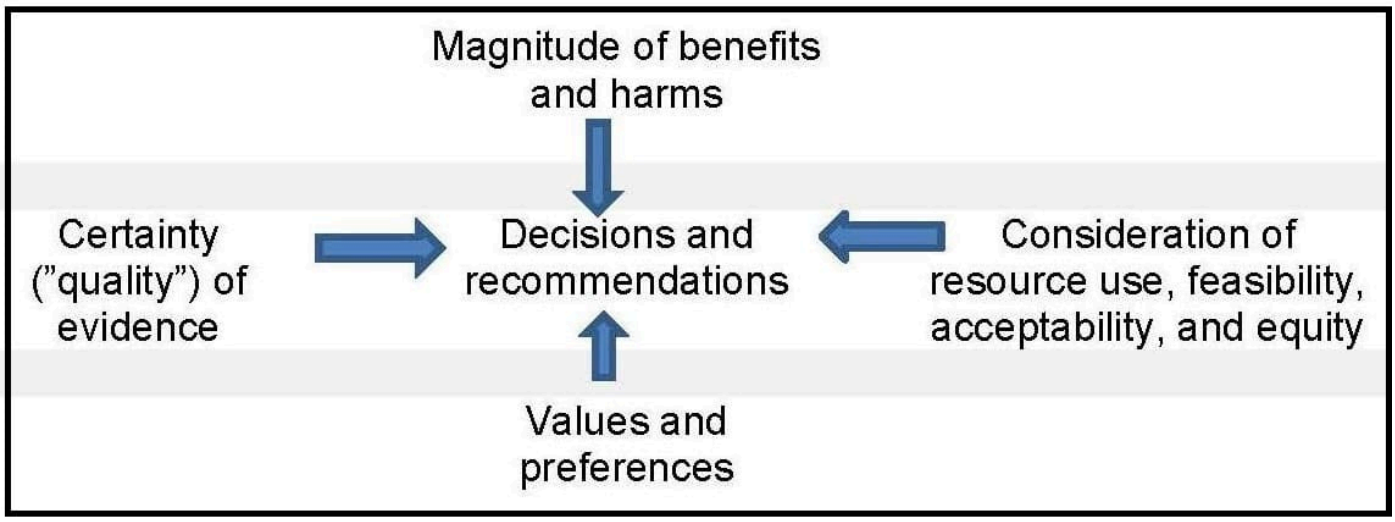


Figure 1.4: Factors affecting decision making according to GRADE27 — Grading of Recommendations assessment, Development, and Evaluation. Redrawn from Djulbegovic and Guyatt (2017).

Recognition of the limitations of the initial traditional EBM hierarchy of evidence led to the concept of a systematic review, now widely used to inform nutrition decisions. A systematic review is the application of scientific strategies to produce comprehensive and reproducible summaries of the relevant scientific literature through the systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific topic (Yetley et al., 2017a). Systematic reviews aim to reduce bias and random error, and provide clarification of the strength and nature of all of the evidence in terms of the quality of research studies, the consistency of the effect, and the evidence of causality. These attributes are particularly useful when there is controversy or conflicting results across the studies (Yetley et al., 2017a).

There are five steps in a systematic review; these are itemized in Box 1.3. Some of their advantages and disadvantages are

summarized in Yetley et al. (2017a)

Box 1.3 Steps in a systematic review

- Prepare the topic — refine questions and develop an analytic framework
- Search for and select studies — identify eligibility criteria, search for relevant studies, and select evidence for inclusion
- Abstract data — extract evidence from studies and construct evidence tables
- Analyze and synthesize data — critically assess quality of studies using a prespecified method, assess applicability of studies, apply qualitative methods, apply quantitative methods such as meta-analysis, and rate the strength of a body of evidence
- Present findings

From Lau in NASEM (2018)

Systematic reviews that address nutrition questions present some unique challenges. Approaches that can be used to address some of these challenges are summarized in Table 1.2 The outcome of a systematic review is an evidence-based review (EBR) which may include quantitative processes such as meta-analyses to analyze and synthesize the data across the studies. However, combining the results of individual studies can lead to misleading conclusions unless the tools described above are applied to ensure the inclusion of the candidate studies is appropriate and they are of high quality. Tools are also available for assessing the overall quality of the evidence generated from a systematic review and meta-analyses. Examples are AMSTAR 2007 and AMSTAR 2 which are available for the conduct , reporting, and subsequent meta-analyses of systematic reviews based on RCTs and non-randomized studies, respectively. For details see AMSTAR. In addition, several Web-based collaborative systematic review tools are available (e.g., SRDR).

Table 1.2 Applying systemmatic reviews to nutrition questions: approaches to the challenges. Data from Brannon (2014).

Challenge	Approach
Baseline exposure	Unlike drug exposure, most persons have some level of diet endogenous synthesis in the case of vitamin D, information that any related uncertainties can be factored into data inte
Nutrient status	The nutrient status of an individual or population can affect
Chemical form of the nutrient or dietary substance	If nutrients occur in multiple forms, the forms may differ in for appropriate data interpretation.
Factors that influence bioavailability	Depending upon the nutrient or dietary substance, influence such as pregnancy may affect the utilization of the nutrient. Capturing such information allows these influences to be fa
Multiple and interrelated biological functions of a nutrient or dietary substance	Biological functions need to be understood in order to ensu
Nature of nutrient or dietary substance intervention	Food-based interventions require detailed documentation c
Uncertainties in assessing dose-response relationships	Specific documentation of measurement and assay procedu

Risk of bias tools are also available for systematic reviews, depending on whether the studies included are randomized or non-randomized. Examples include ROBINS for RCTs and ROBIS that can assess risk of bias in both randomized and nonrandomized studies.

Evidence-grading systems (Figure 1.4) are also used in systematic reviews. Many use GRADE (Guyatt et al., 2011) which uses evidence summaries to systematically grade the evidence as high, moderate, low, or very low for a series of outcomes.

For all systematic reviews, it is important to separate the tasks, with a systematic review team that is separate from the expert group responsible for reviewing the evidence and interpreting the results. Some examples of the misuse of meta-analysis which has led to misleading conclusions can be found in Barnard et al.(2017). Guidelines and guidance to avoid some of the limitations highlighted are available in Dekkers et al. (2019).

Through the Guideline Development Groups (GDGs) at WHO, systematic reviews are now used to inform the scientific judgment

needed for sound evidence-based public health nutrition. The process is used to establish nutrient reference values (NRVs), food-based dietary guidelines, and clinical or public health practice guidelines in dietetics and nutrition.

This page titled 1.2: Nutritional assessment systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1.3: Nutritional assessment methods

Historically, nutritional assessment systems have focused on methods to characterize each stage in the development of a nutritional deficiency state. The methods were based on a series of dietary, laboratory-based biomarkers, anthropometric, and clinical observations used either alone or, more effectively, in combination.

Today, these same methods are used in nutritional assessment systems for a wide range of clinical and public health applications. For example, many low and middle-income countries are now impacted by a triple burden of malnutrition, where undernutrition, multiple micronutrient deficiencies, and overnutrition co-exist. Hence, nutritional assessment systems are now applied to define multiple levels of nutrient status and not just the level associated with a nutrient deficiency state. Such levels may be associated with the maintenance of health, or with reduction in the risk of chronic disease; sometimes, levels leading to specific health hazards or toxic effects are also defined (Combs, 1996).

There is now increasing emphasis on the use of new functional tests to determine these multiple levels of nutrient status. Examples include functional tests that measure immune function, muscle strength, glucose metabolism, nerve function, work capacity, oxidative stress, and genomic stability(Lukaski and Penland, 1996; Mayne, 2003; Russell, 2015; Fenech, 2003).

The correct interpretation of the results of nutritional assessment methods requires consideration of other factors in addition to diet and nutrition. These may often include socioeconomic status, cultural practices, and health and vital statistics, which collectively are sometimes termed “ecological factors”; see Section 1.3.5. When assessing the risk of acquiring a chronic disease, environmental and genetic factors are also important (Yetley et al., 2017a).

1.3.1 Dietary methods

Dietary assessment methods provide data used to describe exposure to food and nutrient intakes as well as information on food behaviors and eating patterns that cannot be obtained by any other method. The data obtained have multiple uses for supporting health and preventing disease. For example, health professionals use dietary data for dietary counseling and education and for designing healthy diets for hospitals, schools, long-term care facilities and prisons. At the population level, national food consumption surveys can generate information on nutrient adequacy within a country, identify population groups at risk, and develop nutrition intervention programs. Dietary data can also be used by researchers to study relationships between diet and disease, and for formulating nutrition policy such as food-based dietary guidelines (Murphy et al., 2016).

It is important to recognize that nutrient inadequacies may arise from a primary deficiency (low levels in the diet) or because of a secondary deficiency. In the latter case, dietary intakes may appear to meet nutritional needs, but conditioning factors (such as certain drugs, dietary components, or disease states) interfere with the ingestion, absorption, transport, utilization, or excretion of the nutrient(s).

Several dietary methods are available, the choice depending primarily on both the study objectives and the characteristics of the study group (see Chapter 3 for more details). Recently, many technical improvements have been developed to improve the accuracy of dietary methods. These include the use of digital photographs of food portions displayed on a cell-phone or a computer tablet, or image-based methods utilizing video cameras, some wearable. Some of these methods rely on active image capture by users, and others on passive image capture whereby pictures are taken automatically. Under development are wearable camera devices which objectively measure diet without relying on user-reported food intake(Boushy et al., 2017). Several on-line dietary assessment tools are also available, all of which standardize interview protocols and data entry: they can be interviewer- or self-administered (Cade, 2017); see Chapter 3 for more details.

Readers are advised to consult Intake — a Center for Dietary assessment that provides technical assistance for the planning, collection, analysis and use of dietary data. Examples of their available publications are presented in Box 1.4. In addition, recommendations for collecting, analyzing, and interpreting dietary data to inform dietary guidance and public health policy are

also available; see Murphy et al., (2016) and Subar et al., (2015) for more details.

Box 1.4 Examples of publications by intake.org

- Considerations for the selection of Portion Size Estimation Methods for Use in Quantitative 24-Hour Dietary Recall Surveys in low- and Middle-Income Countries. (Vossenaar et al. 2020)
- Estimating Usual Intakes from Dietary Surveys: Methodologic Challenges, Analysis Approaches, and Recommendations for low- and Middle-Income Countries. (Tooze, 2020)
- Guidance for the Development of Food Photographs for Portion Size Estimation in Quantitative 24-Hour Dietary Recall Surveys in low- and Middle-Income Countries. (Vossenaar et al. 2020)
- CSDietary Software Program .CSDietary HarvestPlus, SerPro S.A. (CSDietary, 2020)
- Intake Survey Guidance Document: Estimating Usual Intakes from Dietary Surveys — Methodologic Challenges, Analysis Approaches, and Recommendations for LMICs (Tooze, 2020)
- Dietary Survey Protocol Template: An Outline to Assist with the Development of a Protocol for a Quantitative 24-Hour Dietary Recall Survey in a low- or Middle-Income Country. (Dietary Recall, 2020)
- An Overview of the Main Pre-Survey Tasks Required for Large-Scale Quantitative 24-Hour Recall Dietary Surveys in LMICs (Vossenaar et al., 2020)

Data on knowledge, attitudes and practices, and reported food-related behaviors are also collected. Historically, this has involved observing the participants, as well as in-depth interviews and focus groups — approaches based on ethnological and anthropological techniques. Today, e-health (based on the internet) and m-health (based on mobile phones) communication technologies are also being used to collect these data, as noted earlier (Olson, 2016). All these methods are particularly useful when designing and evaluating nutrition interventions.

Often, information on the proportion of the population “at risk” of inadequate intakes of nutrients is required. Such information can be used to ascertain whether assessment using more invasive methods based on nutritional biomarkers are warranted in a specific population or subgroup.

1.3.2 Laboratory Methods

Laboratory methods are used to measure nutritional biomarkers which are used to describe status, function, risk of disease, and response to treatment. They can also be used to describe exposure to certain foods or nutrients, when they are termed “dietary biomarkers”. Most useful are nutritional biomarkers that distinguish deficiency, adequacy and toxicity, and which assess aspects of physiological function and/or current or future health. However, it must be recognized that a nutritional biomarker may not be equally useful across different applications or life-stage groups where the critical function of the nutrient or the risk of disease may be different (Yetley et al., 2017b).

The Biomarkers of Nutrition and Development (BOND) program (Raiten and Combs, 2015) has defined a nutritional biomarker as:

> “a biological characteristic that can be objectively measured and evaluated as an indicator of normal biological or pathogenic processes, and/or as an indicator of responses to nutrition interventions”.

Nutritional biomarkers can be measurements based on biological tissues and fluids, on physiological or behavioral functions and, more recently, on metabolic and genetic data that in turn influence health, well-being, and risk of disease. Yetley and colleagues (2017b) have highlighted the difference between risk biomarkers and surrogate biomarkers. A risk biomarker is defined by the Institute of Medicine (2010) as a biomarker that indicates a component of an individual’s level of risk of developing a disease or level of risk of developing complications of a disease. As an example, metabolomics is being used to investigate potential risk biomarkers of pre-diabetes that are distinct from the known diabetes risk indicators (glycosylated hemoglobin levels, fasting glucose, and insulin) (Wang-Sattler et al., 2012).

BOND classified nutritional biomarkers into three groups shown in Box 1.5,

Box 1.5. Classification of nutritional biomarkers

- Biomarkers of “exposure”: food or nutrient intakes; dietary patterns; supplement usage. Assessed by:
  - Traditional dietary assessment methods
  - Dietary biomarkers: indirect measures of nutrient exposure
- Biomarkers of “status”: body fluids (serum, erythrocytes, leucocytes, urine, breast milk); tissues (hair, nails)
- Biomarkers of “function": measure the extent of the functional consequences of a nutrient deficiency.
  - Functional biochemical: enzyme stimulation assays; abnormal metabolites; DNA damage. These biomarkers serve as early biomarkers of subclinical deficiencies.
  - Functional physiological/behavioral: more directly related to health status or disease such as vision, growth, immune function, taste acuity, cognition, depression. These biomarkers impact on clinical and health outcomes.

In summary:

Exposure → Status → Function → Outcomes

based on the assumption that an intake-response relationship exists between the biomarkers of exposure (i.e., nutrient intake) and the biomarkers of status and function. Functional physiological and behavioral biomarkers are more directly related to health status and disease than are the functional biochemical biomarkers shown in Box 1.5. Disturbances in these functional physiological and behavioral biomarkers are generally associated with more prolonged and severe nutrient deficiency states, and are often affected by social and environmental factors so their sensitivity and specificity are low. In general, functional physiological tests (with the exception of physical growth) are not suitable for large-scale nutrition surveys: they are often too invasive, they may require elaborate equipment, and the results tend to be difficult to interpret because of the lack of cutoff points. Details of functional physiological or behavioral tests dependent on specific nutrients are summarized in Chapters 16–25.

The growing prevalence of chronic diseases has led to investigations to identify biomarkers that can be used as substitutes for chronic disease outcomes (Yetley et al., 2017b). Chronic disease events are characterized by long developmental times, and are multifactorial in nature with challenges in differentiating between casual and associative relations (Yetley et al., 2017b). To qualify as a biomarker that is intended to substitute for a clinical endpoint, the biomarker must be on the major causal pathway between an intervention (e.g., diet or dietary component) and the outcome of interest (e.g., chronic disease). Such biomarkers are termed “surrogate” biomarkers; only a few such biomarkers have been identified for chronic disease. Examples of well accepted surrogate biomarkers are blood pressure within the pathway of sodium intake and cardiovascular disease (CVD) and low density lipoprotein-cholesterol (LDL) concentration within a saturated fat and CVD pathway; see Yetley et al (2017b) for more details.

Increasingly, it is recognized that a single biomarker may not reflect exclusively the nutritional status of that single nutrient, but instead be reflective of several nutrients, their interactions, and metabolism. This has led to the development of “all-in-one” instrument platforms that conduct multiple micronutrient tests in a single sample aliquot, as noted earlier. A 7-plex microarray immunoassay has been developed for ferritin, soluble transferrin receptor, retinol binding protein, thyroglobulin, malarial antigenemia and inflammation status biomarkers (Brindle et al., 2019), which has subsequently been applied to dried blood spot matrices (Brindle et al., 2019). Comparisons with reference-type assays indicate that with some improvements in accuracy and precision, these multiplex instrument platforms could be useful tools for assessing multiple micronutrient biomarkers in national micronutrient surveys in low resource settings (Esmaeili et al., 2019). Readers are advised to consult the Micronutrient Survey Manual and Toolkit developed by the U.S. Centers for Disease Control and Prevention (CDC) for details on planning, implementation, analysis, reporting, dissemination and the use of data generated from a national cross-sectional micronutrient survey. For details, see (CDC, 2020).

1.3.3 Anthropometric methods

Anthropometric methods involve measurements of the physical dimensions and gross composition of the body (WHO, 1995). The measurements vary with age (and sometimes with sex and race) and degree of nutrition, and they are particularly useful in circumstances where chronic imbalances of protein and energy are likely to have occurred. Such disturbances modify the patterns of physical growth and the relative proportions of body tissues such as fat, muscle, and total body water.

In some cases, anthropometric measurements can detect moderate and severe degrees of malnutrition, but cannot identify specific nutrient deficiency states. The measurements provide information on past nutritional history, which cannot be obtained with equal confidence using other assessment techniques.

Anthropometry is used in both clinical and public health settings to identify the increasing burden of both under- and over-nutrition that now co-exist, especially in low- and middle-income countries. Measurements can be performed relatively quickly, easily, and reliably using portable equipment, provided standardized methods and calibrated equipment are used (Chapters 10 and 11). To aid in their interpretation, the raw measurements are generally expressed as an index, such as height-for-age (See Section 1.4).

Standardized methods exist to evaluate anthropometric indices based on Z-scores or percentiles, both calculated in relation to the distribution of the corresponding anthropometric index for the healthy reference population (Section 1.6.1 and Section 1.6.2). Often Z-scores of below –2 or above +2 are used to designate individuals with either unusually low or unusually high anthropometric indices, especially in low income countries. When used in this way, the combination of index and reference limit is termed an "indicator", a term that relates to their use in nutritional assessment, often for public health, or social/medical decision-making (see Chapter 13 for more details).

There is growing concern about the global pandemic of obesity; individuals with obesity are at higher risk of several chronic diseases, including coronary heart disease, diabetes, and hypertension. Consequently, numerous investigators have compared the usefulness of anthropometric variables such as body mass index (weight,kg) / (height, m)^2 (BMI) and waist circumference as surrogate measures of obesity. In a meta-analysis of studies with at least a 12mos follow-up, Seo et al. (2017) concluded that waist circumference was a better predictor for diabetes than BMI (> 30) in women than men and for all ages > 60y, whereas neither BMI > 30, nor waist circumference > 102cm (for men), > 88cm (for women) were significant predictors of hypertension.

1.3.4 Clinical methods

A medical history and a physical examination are the clinical methods used to detect signs, (observations made by a qualified examiner) and symptoms (manifestations reported by the patient) associated with malnutrition or risk of chronic disease. The latter is defined by IOM (2010) as a culmination of a series of pathogenic processes in response to internal or external stimuli over time that results in a clinical diagnosis/ailment and health outcomes; examples include diabetes, cancer, coronary heart disease, stroke, and arthritis. The signs and symptoms may be nonspecific and develop only during the advanced stages of a nutrient deficiency (or excess) or chronic disease; for this reason, their diagnosis should not rely exclusively on clinical methods. It is obviously desirable to have the capacity to detect marginal nutrient deficiencies and risk of chronic disease before a clinical syndrome develops.

Several laboratory-based biomarkers exist to assess an individual's level of risk of developing a disease and as substitutes for chronic disease outcomes; they are often included as an adjunct to clinical assessment. Examples include serum ferritin for risk of iron deficiency anemia, glycosylated hemoglobin (HbA1c) for risk of diabetes, and alterations in bone mineral density for changes in fracture risk. Examples of surrogate biomarkers intended to substitute for chronic disease outcomes include LDL cholesterol instead of the true clinical outcome CVD and blood pressure for cardiovascular disease, as noted earlier (Yetley et al., 2017b).

1.3.5 Ecological factors

Increasingly, nutritional assessment methods include the collection of information on a variety of other factors known to influence the nutritional status of individuals or populations. This increase has stemmed, in part, from the the United Nations Children's Fund (UNICEF) conceptual framework for the causes of childhood malnutrition shown in Figure 1.5, and the increasing focus on studies of diet and chronic disease (Yetley et al., 2017a).

The UNICEF framework highlights that child malnutrition is the outcome of a complex causal process involving not just the immediate determinants such as inadequate dietary intake and poor care, but also the underlying and basic enabling determinants depicted in Figure 1.5. As a consequence, several variables associated with the underlying and enabling determinants of child malnutrition are included in nutritional assessment systems, including in the Demographic Health Surveys conducted in low- and middle-income countries. Variables addressing the underlying determinants include household composition, education, literacy, ethnicity, religion, income, employment, women's empowerment, material resources, water supply, household sanitation,

and hygiene (i.e, WASH) and access to health and agricultural services, as well as land ownership and other information.

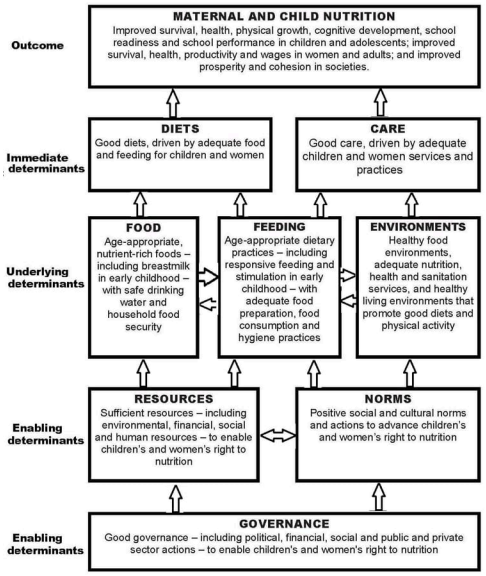


Figure 1.5. A framework for the prevention of malnutrition in all its forms. Redrawn from: UNICEF NUTRITION STRATEGY 2020–2030: UNICEF Conceptual Framework on the Determinants of Maternal and Child Nutrition (2020).

Additional data on food prices, the adequacy of food preparation equipment, the degree of food reserves, cash-earning opportunities, and the percentage of the household income spent on certain foods such as animal foods, fruits, and vegetables can also be collected, if appropriate.

Data on health and vital statistics may also be obtained, as may information on the percentage of the population with ready access to a good source of drinking water, the proportion of children immunized against measles, the proportion of infants born with a low birth weight, the percentage of mothers practicing exclusive breastfeeding up to six months, and and age- and cause-specific mortality rates.

Some of these non-nutritional variables are strongly related to malnutrition and can be used to identify at-risk individuals during surveillance studies. For example, Morley (1973) identified birth order over seven, breakdown of marriage, death of either parent, and episodes of infectious diseases in early life as being important factors in the prediction of West African children who were nutritionally at risk. In a study in the state of Maharashtra in India, Aguayo et al. (2016) reported that after controlling for potential confounding, the most consistent predictors of stunting and poor linear growth in children under 23mos were birthweight and child feeding, women's nutrition and status, and household sanitation and poverty. Women's empowerment has also been shown to significantly influence child nutrition, infant and young child feeding practices, and reproductive health service utilization in some studies (Kabir et al., 2020).

This page titled 1.3: Nutritional assessment methods is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1.4: Nutritional assessment indices and indicators

Raw measurements alone have no meaning unless they are related to, for example, the age or sex of an individual (WHO, 1995). Hence, raw measurements derived from each of the four methods are often (but not always) combined to form "indices." Examples of such combinations include height-for-age, nutrient density (nutrient intake per megajoule), BMI ((weight kg) / (height m)^2), and mean cell volume ((hematocrit) / (red blood cell count)). These indices are all continuous variables. Construction of indices is a necessary step for the interpretation and grouping of measurements collected by nutritional assessment systems, as noted earlier.

Indices are often evaluated in clinical and public health settings by comparison with predetermined reference limits or cutoff points (Section 1.6). Reference limits in anthropometry in low income countries are often defined by Z-scores below –2, as noted earlier. For example, children aged 6–60mos with a height-for-age Z-score < –2 are referred to as "stunted". When used in this way, the index (height-for-age) and the associated reference limit (i.e., < –2 Z-score) are together termed an "indicator", a term used in nutritional assessment, often for public health or social/medical decision-making at the population level.

Several anthropometric indicators have been recommended by the WHO. For example, they define "underweight" as a weight-for-age < –2 Z-score, "stunted" as length/height-for-age < –2 Z-score), and "wasted" as weight-for-length/height < –2 Z-score. In children aged 0–5y, WHO uses a Z-score above +2 for BMI-for-age as an indicator of "overweight", and above +3 as an indicator of obesity (de Onis and Lobstein, 2010). Anthropometric indicators are frequently combined with dietary and micronutrient biomarker indicators for use in public health programs to identify populations at risk; some examples are presented in Table 1.3.

Table 1.3. Examples of dietary, anthropometric, laboratory, and clinical indicators and their application. EAR, estimated average requi

Nutritional indicator
Dietary indicators
Prevalence of the population with zinc intakes below the estimated average requirement (EAR)
Proportion of children 6–23mos of age who receive foods from 4 or more food groups
Anthropometric indicators
Proportion of children age 6–60mos in the population with mid-upper arm circumference < 115mm
Percentage of children < 5y with length- or height-for-age less than –2.0 SD below the age-specific median of the reference population
Lab. indicators based on micronutrient biomarkers
Percentage of population with serum Zn concentrations below the age/sex/time of day-specific lower cutoff
Percentage of children age 6–71mos in the population with a serum retinol < 0.70µmol/L
Median urinary iodine <20µg/L based on > 300 casual urine samples
Proportion of children (of defined age and sex) with two or more abnormal iron indices (serum ferritin, erythrocyte protoporphyrin, transferrin receptor) plus an abnormal hemoglobin
Clinical indicators
Prevalence of goiter in school-age children ≥ 30%
Prevalence of maternal night blindness ≥ 5%

Indicators should be chosen carefully in relation to both the study objectives and their attributes. They can be used to meet a variety of objectives. For example, if the objective of the program is to evaluate the treatment of malnutrition, then the indicator chosen must have the potential to respond to the specific intervention under study and must relate to the nature and severity of

the malnutrition present. Thus, the same indicators are not appropriate for evaluating the treatment of stunting versus wasting. Further, several factors will affect the magnitude of the expected response of an indicator. These may include the degree of deficiency, age, sex, and physiological state of the target group. Other influencing factors may be the type and duration of the intervention, home diet, the age-specificity of the response, and whether the indicator is homeostatically controlled. A more detailed discussion of the selection criteria for indicators can be found in Habicht et al.(1980), Habicht and Pelletier (1990), and Habicht and Stoltzfus (1997).

This page titled 1.4: Nutritional assessment indices and indicators is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1.5: The design of nutritional assessment systems

The design of the nutritional assessment system is critical if time and resources are to be used effectively. The assessment system used, the type and number of measurements selected, and the indices and indicators derived from these measurements will depend on a variety of factors.

Efforts have increased dramatically in the past decade to improve the content and quality of nutritional assessment systems, especially those involving clinical trials. In 2013, guidelines were published on clinical trial protocols entitled: Standard Protocol Items: Recommendations for International Trials (SPIRIT). This has led to the compulsory preregistration of clinical trials, and often publication of the trial protocols in scientific journals. The SPIRIT checklist consists of 33 recommended items to include in a clinical trial. Chan et al. (2013) provide the rationale, a detailed description, and model example of each item. Discussions on compulsory preregistration of protocols for observational studies are in progress; see Lash and Vandenbroucke (2012).

An additional suggestion to support transparency and reproducibility in clinical trials, and to distinguish data-driven analyses from pre-planned analyses is the publication of a statistical analysis plan before data have been accessed (DeMets et al., 2017; Gamble et al., 2017). Initially, recommendations for a pre-planned statistical analyses plan were compiled only for clinical trials (Gamble et al., 2017), but have since been modified for observational studies by Hiemstra et al. (2019) to include details on the adjustment for possible confounders. Tables of the recommended content of statistical analysis plans for both clinical trials and observational studies are also available in Hiemstra et al. (2019).

1.5.1 Study objectives and ethical issues

The general design of the assessment system, the raw measurements, and, in turn, the indices and indicators derived from these measurements should be dictated by the study objectives. Possible objectives may include:

1. Determining the overall nutritional status of a population or subpopulation
2. Identifying areas, populations, or subpopulations at risk of chronic malnutrition
3. Characterizing the extent and nature of the malnutrition within the population or subpopulation
4. Identifying the possible causes of malnutrition within the population or subpopulation
5. Designing appropriate intervention programs for high-risk populations or subpopulations
6. Monitoring the progress of changing nutritional, health, or socioeconomic influences, including intervention programs
7. Evaluating the efficacy and effectiveness of intervention programs
8. Tracking progress toward the attainment of long-range goals.

The first three objectives can be met by a cross-sectional nutrition survey, often involving all three of the major methods of nutritional assessment. Such surveys, however, are unlikely to provide information on the possible causes of malnutrition (i.e., objective no. 4). The latter can only be achieved through interventions (objectives no. 5 and no. 7) and possibly objective no. 6. An assessment of the possible causes of malnutrition is a necessary prerequisite when implementing nutrition intervention programs.

In some circumstances, the objective may be to identify only those individuals at risk of malnutrition and who require intervention (i.e., objective no. 5). To achieve this objective, a screening system is required that uses simple and cheap measurements and reflects both past and present nutritional status.

Ethical issues

Formal guidelines on the general conduct of biomedical research is contained in the declaration of Helsinki on Ethics and Epidemiology, published by the Council for International Organization of Medical Sciences (CIOMS, 2016). Ethical approval from the appropriate human ethics committees in the countries involved in the research study must be obtained by the principal investigators before work begins. The basic guidelines for research on human subjects must be followed. As an example, sections of the regulations of the U.S. Department of Health and Human Services (2021) are shown in Box 1.6.

Box 1.6: Some possible guidelines for research on human subjects

- **Risks** to subjects are minimized and proportional to the anticipated benefits and knowledge.
- **Data** are monitored to ensure safety of subjects.
- **Selection of subjects** is equitable.
- **Vulnerable subjects**, if included, are covered by additional safeguards.
- **Informed consent** is obtained from the subjects.
- **Confidentiality** is adequately protected.

From DHHS (2021)

A more detailed discussion of the main ethical issues when planning an application for research ethical approval is available in Gelling (2016).

Informed consent must be obtained from the participants or their principal caregivers in all studies. When securing informed consent, the investigator should also:

- Disclose details of the nature and procedures of the study
- Clearly state the associated potential risks and benefits
- Confirm that participation in the research is voluntary
- Confirm that participants are free to withdraw from the study at any time
- Explain how the results relating to individual participants will be kept confidential
- Describe the procedures that provide answers to any questions and further information about the study.

With the increasing reliance on randomized clinical trials (RCT) to inform evidence-based practice, there have been coordinated attempts to standardize reporting and to register information about trials for consistency and transparency. This has led to the publication of the Consolidated Standards of Reporting Trials (CONSORT). The CONSORT guidelines specify details that should be well-defined in every RCT, and many journals now require these guidelines to be addressed as a condition of publication. The first CONSORT guidelines were published in 2001 and were revised in 2010 and updated frequently. See: Moher et al. (2012).

In 2004, members of the International Committee of Medical Journal Editors (ICMJE) agreed to require registration of any RCT submitted for review and possible publication (DeAngelis et al., 2004). Several registries have been developed which meet the following ICMJE criteria. These include the registry should be accessible to the public; there should be no charge for registration; open to all interested registrants, managed by a nonprofit organization, and have a means for verifying the validity of the registered information (Elliot, 2007).

Standards have now been developed to Strengthen Observational Studies in Epidemiology (STROBE). The STROBE guidelines include 18 items common to three study designs, with four additional items specific for cohort, case-control, or cross-sectional studies (von Elm et al., 2014).Registration of protocols for observational studies may be mandatory in the future (Williams et al., 2010).

Standards have also been developed for reporting qualitative research. Two reporting standards are often used — the Consolidated Criteria for Reporting Qualitative Research (COREQ) (32-item checklist) (Tong et al., 2007) and the Standards for Reporting Qualitative Research (SRQR) (21-item checklist (O'Brien et al., 2014). Their use can assist researchers to report important aspects of the research team, study methods, context of the study, findings, analysis and interpretations.

1.5.2 Choosing the study participants and the sampling protocol

Nutritional assessment systems often target a large population — perhaps that of a city, province, or country. That population is best referred to as the “target population”. To ensure that the chosen target population has demographic and clinical characteristics of relevance to the question of interest, a specific set of inclusion criteria should be defined. However, for practical reasons, only a limited number of individuals within the target population can actually be studied. Hence, these individuals must be chosen carefully to ensure the results can be used to infer information about the target population. This can be achieved by defining a set of exclusion criteria to eliminate individuals who it would be unethical or inappropriate to study; as few exclusion criteria as possible should be specified. The technique of selecting a sample representative of the target population and of a size adequate

for achieving the primary study objectives, requires the assistance of a statistician; only a very brief review is provided here.

A major factor influencing the choice of the sampling protocol is the availability of a sampling frame. Additional factors include time, resources, and logistical constraints. The sampling frame is usually a comprehensive list of all the individuals in the population from which the sample is to be chosen. In some circumstances, the sampling frame may consist of a list of districts, villages, institutions, school or households, termed "sampling units" rather than individuals *per se*.

When a sampling frame is not available, nonprobability sampling methods must be used. Three nonprobabilty sampling methods are available: consecutive sampling, convenience sampling, and quota sampling, each of which is described briefly in Box 1.7. Note that the use of nonprobability sampling methods produces samples that may not be representative of the target population and hence may lead to systematic bias: such methods should be fully documented.

Box 1.7 Nonprobability sampling protocols

- **Consecutive sampling** is the best of the nonprobability techniques. It is used, for example, in clinical research, when it is feasible to recruit all available patients who meet the selection criteria, over a time period that is long enough to avoid seasonal factors or other changes over time.
- **Convenience sampling** involves taking individuals into the study who happen to be available at the time of data collection and who consent to participate. This approach is used when the population to be sampled is relatively homogeneous
- **Quota sampling** involves dividing the target population into a number of different categories based on age, ownership of land, or occupations etc, and taking a certain number of consenting individuals from each category into the final sample.

Several possible sources of bias can occur when nonprobability sampling is used, as shown by the three domains (nos 1–3) depicted in Figure 1.3. Some specific examples include the following:

- Ignoring people who do not respond to an initial approach to include them in the study — the non-response bias. People who refuse to take part may have characteristics that differ markedly from those of the respondents.
- Studying only volunteers who are often unrepresentative of most of the population.
- Sampling only those persons attending a clinic, school, or health center, and neglecting to include non-attendees.
- Collecting data at only one time of the year, which may introduce a seasonal bias.
- Selecting participants who are accessible by road introduces a "tarmac" bias. Areas accessible by road are likely to be systematically different from those that are more difficult to reach.

It is essential to fully document the characteristics of the sample and to identify the probable direction and magnitude of the bias that arises from the adopted sample protocol and nonresponse rate. Extrapolating the results from a nonprobability sample to the target population is risky and should be avoided.

Every attempt should be made to compile some type of sampling frame, or to use one that already exists, so that probability sampling can be used(Lemeshow et al., 1990). Probability sampling is the recommended method for obtaining a representative sample with minimum bias.

In settings where maps and census data are out of date or non-existent such as in poor urban environments, creating a sampling frame to select a representative sample is particularly challenging. Investigators working in urban slums in four low-income countries have described a method for creating a spatially-referenced sampling frame consisting of a census of all households in a slum from which a spatially-regulated representative sample can be generated; see the Improving Health in Slums Collaborative (2019) for more details.

Several probability sampling methods exist: simple random sampling, systematic sampling, stratified random sampling, cluster sampling, and multistage sampling. Every effort must be made to minimize the number of nonrespondents so that the generalizability (i.e., external validity) of the study is not compromised. The level of nonresponse that will compromise the generalizability of the study depends on the nature of the research question and on the reasons for not responding. Strategies exist for minimizing refusal to participate in the study; see Hulley et al. (2013) for more details.

Of the probability sampling methods, three are described in Box 1.8; further details can be found in Varkevisser et al. (1993).

Boxx 1.8 Probability sampling protocols

- Simple random sampling involves drawing a random sample from a listing of all the people in the target population.
- Stratified random sampling divides the target population into a number of subgroups or strata (e.g., urban and rural populations, different ethnic groups, various geographical areas, or administrative regions). A separate random sample is then drawn from each of the strata. The stratified subsamples can be weighted to draw disproportionately from subgroups that are less common in the population but of special interest.
- random sampling requires defining a number of levels of sampling, from each of which is drawn a random sample.

Cluster sampling requires defining a random sample of natural groupings (clusters) of individuals in the population. This method is used when the population is widely spaced and it is difficult to compile a sampling frame, and thus sample from all its elements. Statistical analysis must take clustering into account because cluster sampling tends to result in more homogeneous groups for the variables of interest in the population.

Stratified sampling results in a sample that is not necessary representative of the actual population. The imbalance can be corrected, however, by weighting, allowing the results to be generalized to the target population. Alternatively, a sampling strategy, termed proportional stratification, can be used to adjust the sampling before selecting the sample, provided information on the size of the sampling units is available. This approach simplifies the data analysis and also ensures that subjects from larger communities have a proportionately greater chance of being selected than do subjects from smaller communities.

Multistage random sampling is frequently used in national nutrition surveys. It typically involves sampling at four stages: at the provincial or similar level (stage one), at the district level (stage two), at the level of communities in each selected district (stage three), and at the household level in each chosen community (stage four). A random sample must be drawn at each stage. The U.S. NHANES III, the U.K. Diet and Nutrition surveys, and the New Zealand and Australian national nutrition surveys all used a combination of stratified and multistage random sampling techniques to obtain a sample representative of the civilian non-institutionalized populations of these countries.

As can be seen, each probability sampling protocol involves a random selection procedure to ensure that each sampling unit (often the individual) has an equal probability of being sampled. Random selection can be achieved by using a table of random numbers, a computer program that generates random numbers, or a lottery method; each of these procedures is described in Varkisser et al. (1993).

1.5.3 Calculating sample size

The appropriate sample size for a particular nutritional assessment project should be estimated early in the process of developing the project design so that, if necessary, modifications to the design can be made. The number of participants required will depend on the study objective, the nature and scope of the study, and the “effect size” — the magnitude of the expected change or difference sought. The estimate obtained from the sample size calculation represents the planned number of individuals with data at outcome, and not the number who should be enrolled. The investigator should always plan for dropouts and individuals with missing data.

The first step in the process of estimating the sample size is restating the research hypothesis to one that proposes no difference between the groups that are being compared. This restatement is called the “null” hypothesis. Next, the “alternative” hypothesis should be stated, which, if one-sided, specifies the actual magnitude of the expected “effect size” and the direction of the difference between the predictor and outcome variable. In most circumstances, however, a two-sided alternative hypothesis is stated, in which case only the effect size is specified and not the direction.

The second step in the estimation of sample size is the selection of a reasonable effect size (and variability, if necessary). As noted earlier, this is rarely known, so instead both the effect size and variability must be estimated based on prior studies in the literature, or selected on the basis of the smallest effect size that would be considered clinically meaningful. Sometimes a small pilot study is conducted to estimate the variability (s^2) of the variable. When the outcome variable is the change of a continuous measurement (e.g., change in a child's length during the study), the s^2 used should be the variance of this change.

The third step involves setting both α and β. The probability of committing a type 1 error (rejecting the null hypothesis when it is actually true) is defined as α. Another widely used name for α is the level of significance. It is often set at 0.05, when it represents a 95% assurance that a significant result will not be achieved when it should not (i.e., the null hypothesis will not be rejected). If

a one-tailed alternative hypothesis has been set, then a one-tailed α should be used; otherwise, use a two-tailed α.

The probability of committing a type II error (i.e.,failing to reject the null hypothesis when it is actually false) is defined as "β", and is often set at 0.20, indicating that the investigator is willing to accept a 20% chance of missing an association of the specified effect size if it exists. The quantity 1–β is called the power, and when set at 0.80 implies there is a 80% chance of finding an association of that size or greater when it really exists.

The final step involves selecting the appropriate procedure for estimating the sample size. Two different procedures can be used depending on how the effect size is specified. Frequently, the objective is to determine the sample size to detect differences in the **Proportion of individuals** in two groups. For example, the proportion of male infants age 9mos who develop anemia while being treated with iron supplements (Hemoglobin < 110g/L) is to be compared to the proportion who develop anemia while taking a placebo. The procedure is two-sided, allowing for the possibility that the placebo is more effective than the supplement! Note that the effect size is the difference in the projected proportions in the two groups and that the size of that differences critically controls the required sample size. See, Sample size calculator - two proportions.

In a cohort or experimental study, the effect size is the difference between P_1, the proportion of individuals expected to have the outcome in one group and P_2, the proportion expected in the other group. Again, this required effect size must be specified, along with α and β to calculate the required sample size.

In contrast, in a case-control study, P_1 represents the proportion of cases expected to have a particular dichotomous predictor variable (i.e., the prevalence of that predictor), and P_2 represents the proportion of controls who are expected to have the dichotomous predictor.

For examples when the effect size is specified in terms of relative risk or odds ratio (OR), see Browner et al. in Chapter 6 in Hulley et al. (2013).

Alternatively, the objective may be to calculate an appropriate sample size to detect if the **mean value of a continuous variable** in one group differs significantly from the mean of another group. For example, the objective might be to examine the mean HAZ-score of city childen aged 5y with their rural counterparts aged 5y. The sample size procedure assumes that the distribution of the variable in each of the two groups will be approximately normal. However, the method is statistically robust, and can be used in most situations with more than about 40 individuals in each group. Note that in this cases the effect size is the numerical difference in the means of the two groups and that the group variance must also be defined. See,Sample size calculator - two means. However,this sample size calculator cannot be used for studies involving more than two groups, when more sophisticated procedures are needed to determine the sample size.

A practical guide to calculating the sample size is published by WHO (Lwanga and Lemeshow, 1991). The WHO guide provides tables of minimum sample size for various study conditions (e.g., studies involving population proportion, odds ratio, relative risk, and incidence rate), but the tables are only valid when the sample is selected in a statistically random manner. For each situation in which sample size is to be determined, the information needed is specified and at least one illustrative example is given. In practice, the final sample size may be constrained by cost and logistical considerations.

1.5.4 Collecting the data

Increasingly, digital tablets rather than paper-based forms are used for data collection. Their use reduces the risk of transcription errors, and can protect data security through encryption. The transport and storage of multiple paper forms is eliminated and costs can be reduced by the elimination of extensive data entry. Several proprietary and open-source software options (e.g., Open Data Kit) are available for data collection. Initially, data are usually collected and stored locally offline, but uploaded on to a secure central data store when internet access is available.

The process of data aquisition, organisation, and storage should be carefully planned in advance, with the objective of facilitating subsequent data handling and analysis and minimising data entry errors — a particular problem with dietary data.

1.5.5 Additional considerations

Of the many additional factors affecting the design of nutritional assessment systems, the acceptability of the method, respondent burden, equipment and personnel requirements, and field survey and data processing costs are particularly important. The

methods should be acceptable to both the target population and the staff who are performing the measurements. For example, in some settings, drawing venous blood for biochemical determinations such as serum retinol may be unacceptable in infants and children, whereas the collection of breast milk samples may be more acceptable. Similarly, collecting blood specimens in populations with a high prevalence of HIV infections may be perceived to be an unacceptable risk by staff performing the tests.

To reduce the nonresponse rate and avoid bias in the sample selection, the respondent burden should be kept to a minimum. In the U.K. Diet and Nutrition Survey, the seven-day weighed food records were replaced by a four-day estimated food diary, when the rolling program was introduced in 2008 due to concerns about respondent burden (Ashwell et al., 2006). Alternative methods for minimizing the nonresponse rate includes the offering of material rewards and the provision of incentives such as regular medical checkups, feedback information, social visits, and telephone follow-up.

The requirements for equipment and personnel should also be taken into account when designing a nutritional assessment system. Measurements that require elaborate equipment and highly trained technical staff may be impractical in a field survey setting; instead, the measurements selected should be relatively noninvasive and easy to perform accurately and precisely using rugged equipment and unskilled but trained assistants. The ease with which equipment can be transported to the field, maintained, and calibrated must also be considered.

The field survey and data processing costs are also important factors. Increasingly, digital tablet devices are being used for data collection in field surveys rather than paper-based forms. As noted earlier, adoption of this method reduces the risk of transcribing error, protects data security through encryption and reduces the cost of extensive data entry. Several proprietary and open-source software options are available, including Open Data Kit, RedCap, and Survey CTO. Software such as Open Data Kit permits offline data collection, automatic encryption, and the ability to upload all submissions when a data collection devise, such as a notebook, is connected to the internet.

In surveillance systems, the resources available may dictate the number of malnourished individuals who can subsequently be treated in an intervention program. When resources are scarce, the cutoff point for the measurement or test (Section 1.6.3) can be lowered, a practice that simultaneously decreases sensitivity, but increases specificity, as shown in Table 1.7. As a result, more truly malnourished individuals will be missed while at the same time fewer well-nourished individuals are misdiagnosed as malnourished.

This page titled 1.5: The design of nutritional assessment systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1.6: Important characteristics of assessment measures

All assessment measures vary in their validity, sensitivity, specificity, and predictive value; these characteristics, as well as other important attributes, are discussed below,

1.6.1. Validity

Validity is an important concept in the design of nutritional assessment systems. It describes the adequacy with which a measurement or indicator reflects what it is intended to measure. Ideally valid measures are free from random and systematic errors and are both sensitive and specific (Sections 1.6.4; 1.6.5; 1.6.7; 1.6.8).

In dietary assessment, a method that provides a valid reflection of the true "usual nutrient intake" of an individual is often required. Hence, a single weighed food record, although the most accurate dietary assessment method, would not provide a valid assessment of the true "usual nutrient intake" of an individual, but instead provides a measurement of the actual intake of an individual over one day. Similarly, if the biomarker selected reflects "recent" dietary exposure, but the study objective is to assess the total body store of a nutrient, the biomarker is said to be invalid. In the earlier U.S. NHANES I survey, thiamine and riboflavin were analyzed in casual urine samples because it was not practical to collect 24h urine samples. However, the results were not indicative of body stores of thiamine or riboflavin, and hence were considered invalid; the determination of thiamine and riboflavin in casual urine samples were not included in U.S. NHANES II or U.S. NHANES III (Gunter and McQuillan, 1990).

In some circumstances, assessment measures only have "internal" validity, indicating that the results are valid only for the particular group of individuals being studied and cannot be generalized to the universe. In contrast, if the results have "external" validity, or generalizability, then the results are valid when applied to individuals not only in the study but in the wider universe as shown in Figure 1.6. For example, conclusions derived from a study on African Americans may be valid for that particular population (i.e., have internal validity) but cannot be extrapolated to the wider American population. Internal validity is easier to achieve. It is necessary for, but does not guarantee, external validity. External validity requires external quality control of the measurements and judgment about the degree to which the results of a study can be extrapolated to the wider universe. The design of any nutritional assessment system must include consideration of both the internal and external validities of the raw measurements, the indices based on them, and any derived indicators, so that the findings can be interpreted accordingly.

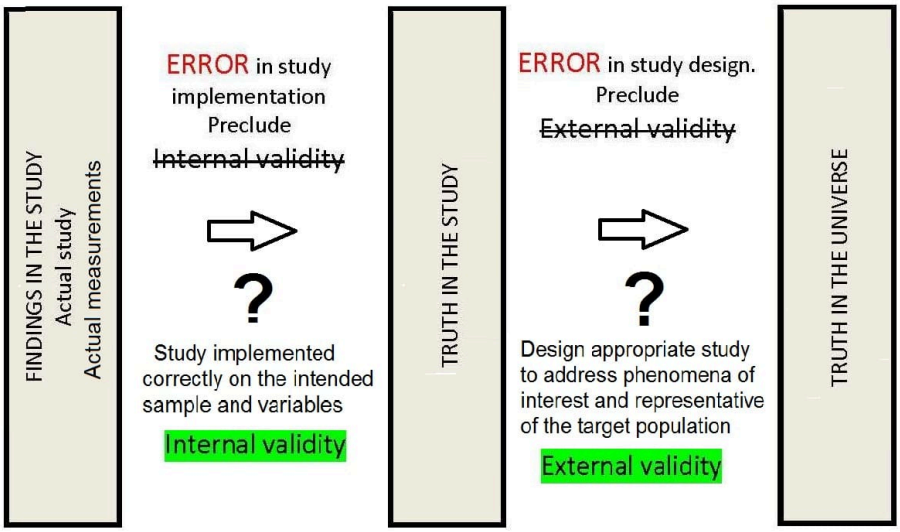


Figure 1.6. External and Internal validity. Redrawn from Hulley et al. (2013)

1.6.2. Reproducibility or precision

The degree to which repeated measurements of the same variable give the same value is a measure of reproducibility —

also referred to as “reliability” or “precision” in anthropometric (Chapter 9) and laboratory assessment (Chapter 15). The measurements can be repeated on the same subject or sample by the same individual (within-observer reproducibility) or different individuals (between-observer reproducibility). Alternatively, the measurements can be assessed within or between instruments. Reproducible measurements yield greater statistical power at a given sample size to estimate mean values and to test hypotheses.

The study design should always include some replicate observations (repeated but independent measurements on the same subject or sample). In this way, the reproducibility of each measurement can be calculated. When the measurements are continuous, the coefficient of variation (CV%) can be calculated:

$$\text{CV } \% = \text{ standard deviation } \times 100\% / \text{ mean}$$

For categorical variables, percent agreement, the interclass correlation coefficient, and the kappa statistic can be used.

In anthropometry, alternative methods are often used to assess the precision of the measurement techniques; these are itemized in Box 1.9, and discussed in Chapter 9. The TEM was calculated for each anthropometric measurement used in the WHO Multicenter Growth Reference Study for the development of the Child Growth Standards; see de Onis et al. (2004).

Box 1.9 Measures of the precision of anthropometric measurements

- Technical Error of the Measurement (TEM)
- Percentage Technical error (% TEM)
- Coefficient of Reliability

The reproducibility of a measurement is a function of the random measurement errors (Section 1.6.4) and, in certain cases, true variability in the measurement that occurs over time. For example, the nutrient intakes of an individual vary over time (within-person variation), and this results in uncertainty in the estimation of usual nutrient intake. This variation characterizes the true “usual intake” of an individual. Unfortunately, within-person variation cannot be distinguished statistically from random measurement errors, irrespective of the design of the nutritional assessment system (see Chapter 6 for more details).

The precision of biochemical measures is similarly a function of random errors that occur during the actual analytical process and within-person biological variation in the biochemical measure. The relative importance of these two sources of uncertainty vary with the different measures. For many modern biochemical measures, the within-person biological variation now exceeds the long-term analytical variation, as shown in Table 1.4.

Table 1.4 Within-person and analytical variance components for some common biochemical measures. Abstracted from Gallagher et al. (1992).

	Coefficient of variation (%)	
Measurement	**Within-person**	**Analytical**
Serum retinol		
Daily	11.3	2.3
Weekly	22.9	2.9
Monthly	25.7	2.8
Serum ascorbic acid		
Daily	15.4	0.0
Weekly	29.1	1.9
Monthly	25.8	5.4
Serum albumin		
Daily	6.5	3.7
Weekly	11.0	1.9

Measurement	Coefficient of variation (%)	
	Within-person	Analytical
Monthly	6.9	8.0

A variety of strategies can be used to minimize random measurement errors and increase the reproducibility of nutritional assessment systems. These strategies were adopted by the WHO Multicenter Growth Reference Study, and are described in de Onis et al. (2004). They included the following:

- Compiling an operations manual that contains specific written guidelines for taking each measurement, to ensure all the techniques are standardized
- Training all the examiners to use the standardized techniques consistently; the latter is especially important in large surveys involving multiple examiners, and in longitudinal studies, where maintaining standardized measurement techniques during the survey is an important issue
- Carefully selecting and standardizing the instruments used for the data collection; in some cases, variability can be reduced by the use of automated instruments
- Refining and standardizing questionnaires and interview protocols, preferably, where feasible, with the use of computer-administered interview protocols; the latter approach is now often used for 24-h recall interviews in national nutrition surveys
- Reducing the effect of random errors from any source by repeating all the measurements, when feasible, or at least on a random subsample.

1.6.3. Accuracy

The term “accuracy” is best used in a restricted statistical sense to describe the extent to which the measurement is close to the true value. It therefore follows that a measurement can be reproducible or precise, but, at the same time, inaccurate — a situation which occurs when there is a systematic bias in the measurement (see Figure 1.7 and Section 1.6.5). The greater the systematic error or bias, the less accurate the measurement. Accurate measurements, however, necessitate high reproducibility, as shown in Figure 1.7. Accuracy is not affected by sample size.

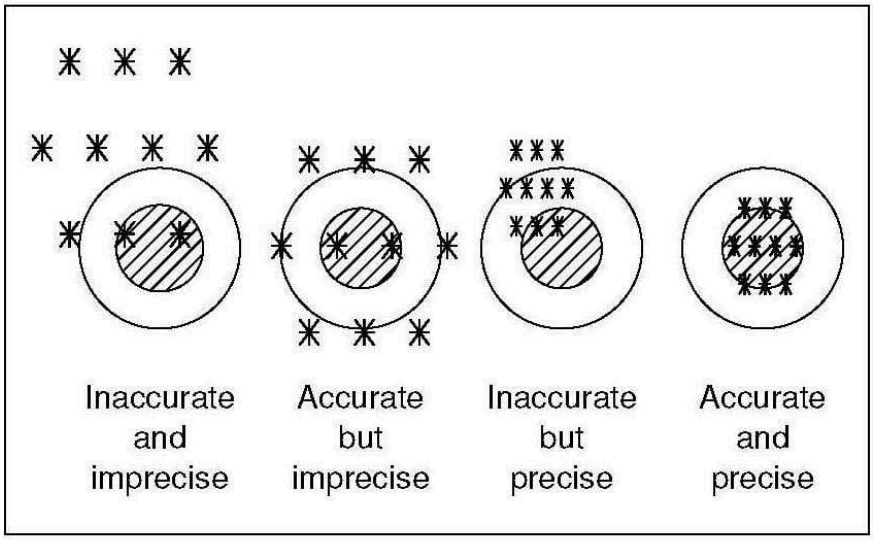


Figure 1.7 Differences between precision and accuracy.

Several approaches exist for assessing the accuracy of a measurement, which vary according to the method being used in the nutritional assessment system. Each approach aims to use a reference measurement undertaken by a technique that is believed to best represent the true value of the characteristic. The reference method is termed a “gold standard”.

Assessing the accuracy of objective measurements of biochemical biomarkers is relatively easy and can be accomplished by using reference materials with certified values for the nutrient of interest, preferably with values that span the concentration range observed in the study; see Chapter 15 for more details. Certified reference materials can be obtained from the U.S. National

Institute of Standards and Technology(e.g., NIST), the U.S. Centers for Disease Control (CDC), the International Atomic Energy Authority (IAEA) in Vienna, the Community Bureau of Reference of the Commission of the European Communities (BCR) in Belgium, and the U.K. National Institute of Biological Standards and Controls (NIBSC).

Table 1.5 Precision and accuracy of measurements.

	Precision or reproducibility	Accuracy
Definition	The degree to which repeated measurements of the same variable give the same value	The degree to which a measurement is close to the true value
Assess by	Comparison among repeated measures	Comparison with certified reference materials, criterion method, or criterion anthropometrist
Value to study	Increases power to detect effects	Increases validity of conclusions
Adversely affected by	Random error contributed by the measurer, the respondent, or the instrument	Systematic error (bias) contributed by: the measurer, the respondent, or the instrument

The control of accuracy in other nutritional assessment methods is more difficult and is discussed in more detail in later chapters. For example, the correct value of any anthropometric measurement is never known with absolute certainty. In the absence of absolute reference standards, the accuracy of anthropometric measurements is assessed by comparing them with those made by a designated criterion anthropometrist (Table 1.5). This approach was used in the WHO Multicenter Growth Reference Study; see de Onis et al. (2004) and Chapter 9 for more details.

Accurate measurements must also be reproducible or precise (Figure 1.7), as noted earlier. Therefore, the same strategies outlined under reproducibility (Section 1.6.2) should be adopted, with the exception of repeating the measurements. Additional strategies that can also be used to enhance accuracy include (a) making unobtrusive measurements, (b) blinding, and (c) calibrating the instruments. Of these strategies, the first two should always be used to help avoid bias where feasible and appropriate. An example of a strategy based on unobtrusive measurements to enhance accuracy in dietary assessment is surreptitious weighing of food portions consumed by the participants in institutional settings such as school lunch programs (Warren et al., 2003). Blinding is used in double-blind clinical trials to ensure that neither the participants nor the researchers know in which group they have been assigned. This strategy, although not ensuring the overall accuracy of the measurements, is practiced to minimize the possibility that the apparent effects of the intervention are due to differential use of other treatments in the intervention and control groups, or to biased judgement of the outcome (Hulley et al., 2013). The third strategy, calibrating the instruments, should always be used when any instruments are involved.

The strategies actually adopted to maximize reproducibility and accuracy will depend on several factors. These may include feasibility and cost considerations, the importance of the variable, and the magnitude of the potential impact of the anticipated degree of inaccuracy on the study conclusions (Hulley et al., 2013).

1.6.4. Random errors

Random errors generate a deviation from the correct result due to chance alone. They lead to measurements that are imprecise in an unpredictable way, resulting in less certain conclusions. They reduce the precision of a measurement by increasing the variability about the mean. They do not influence the mean or median value.

There are three main sources of random error:

- Individual biological variation
- Sampling error
- Measurement error

Individual biological variation may be a major source of error (see Table 1.4). Variability due to time of day may affect both anthropometric measurements (e.g., height) and biochemical measurements (e.g., serum iron and serum zinc). Some nutritional biomarkers also fluctuate in response to medication or meal consumption; for serum zinc, for example, variations in response to meal consumption can be as much as 20% (King et al., 2018).

Sampling may also be a major source of random error. For example, significant sampling errors may be associated with the selection of respondents, who for practical reasons, are usually a small subset of a larger population, or with the collection of a particular type of food (e.g., maize porridge) for nutrient analysis. Such errors will be present even if the sampling is truly random. One way to reduce this error is to increase the sample size (i.e., the number of subjects or the number of maize porridge samples).

Measurements may also generate random errors. During 24-hr dietary recall interviews, for example, a major source of random measurement error may be associated with the measurement of the actual portion size of the foods consumed (Chapter 5). Random measurement errors in anthropometry may also arise from variations during the measurement in the compressibility of the skin by skinfold calipers (Ward and Anderson, 1993) and restless infants when measuring recumbent length.

Random measurement errors can be minimized by using standardized measurement techniques and trained personnel and by employing rigorous analytical quality-control procedures during laboratory analysis. However, such errors can never be entirely eliminated. To be sure, random measurement errors may be generated when the same examiner repeats the measurements (within- or intra-examiner error), or when several different examiners repeat the same measurement (between- or inter-examiner error). Details of the quality control procedures that can be incorporated to minimize sources of measurement error during dietary and biomarker assessment are included in Chapter 5 and 15, respectively.

1.6.5. Systematic errors or bias

Unfortunately, systematic errors may arise in any nutritional assessment method, causing it to become biased. Bias may be defined as a condition that causes a result to depart from the true value in a consistent direction. The errors arising from bias reduce the accuracy of a measurement by altering the mean or median value. They have no effect on the variance and hence do not alter the reproducibility or precision of the measurement(Himes, 1987).

Several types of bias exist, as shown in Figure 1.3. A detailed list of sources of bias that can affect nutrition studies is available in Yetley et al. (2017a), but the principal biases are selection bias and measurement bias. All types of nutritional assessment systems may experience selection bias. It arises when there is a systematic difference between the characteristics of the individuals selected for the study and the characteristics of those who are not, making it impossible to generalize the results to the target population. Selection bias may originate in a variety of ways. Some of these are outlined in Box 1.10.

Box 1.10: Various types of selection bias

- Self-selection bias results from studying only volunteers for a study, who perhaps volunteer because they are unwell.
- Referral bias will be present if cases are recruited through a district health center but controls from an adjacent village.
- nonresponse bias is caused by ignoring people who do not respond to an initial attempt to include them in the study.
- Diagnostic bias arises from selecting subjects for a multicenter case-control study using different diagnostic criteria in different centers.
- Drop-out bias is usually the result of ignoring possible systematic differences between those who fail to complete a study and the remaining participants.

Wherever possible, a strategy should be used to obtain information on people who refuse to participate or subsequently fail to complete the study. This information can then be used to assess whether those who did not participate or dropped out of the study are similar to the participants. If they differ, then a selection bias is present.

Measurement bias can be introduced in a variety of ways. For example:

Biased equipment may over- or underestimate weight or height. Alternatively, skinfold calipers may systematically over- or under-estimate skinfold thickness because of differences in the degree of compression arising from the magnitude of the jaw pressure.

Analytical bias may result from the use of a biochemical method that systematically under- or overestimates the nutrient content of a food or biological specimen. For example, vitamin C may be underestimated because only the reduced form of vitamin C, and not total vitamin C, is measured. Alternatively, if the biopsy specimens of the treatment and control groups are analyzed in different laboratories which produce systematically different results for the same assay, then the assay results will be biased.

Social desirability bias occurs, for example, when respondents underestimate their alcohol consumption in a 24h food recall or record. However, scales can be used to measure the extent of the bias (Robinson et al., 1991), including the Marlowe-Crowne

Social Desirability Scale (Crowne and Marlowe, 1960).

Interviewer bias arises when interviewers differ in the way in which they obtain, record, process, and interpret information. This is a particular problem if different interviewers are assigned different segments of the population, such as different racial or age groups.

Recall bias is a form of measurement bias of critical importance in retrospective case control studies. In such studies, there may be differential recall of information by cases and controls. For example, persons with heart disease will be more likely to recall past exposure to saturated fat than the controls, as saturated fat is widely known to be associated with heart disease. Such a recall bias may exaggerate the degree of effect of association with the exposure or, alternatively, may underestimate the association if the cases are more likely than controls to deny past exposure.

Bias is important as it cannot be removed by subsequent statistical analysis. Consequently, care must be taken to reduce and, if possible, eliminate all sources of bias in the nutritional assessment system by the choice of an appropriate design and careful attention to the equipment and methods selected. Strategies for controlling bias and its potential effect on the measurement of a cause-effect relationship are described in most standard epidemiological texts; see Hulley et al. (2013) for more details. For examples of criteria that can be applied to assess the risk of bias depending on the type of study (i.e., RCT, cohort, case-control, cross-sectional), see Yetley et al. (2017a).

1.6.6. Confounding

Confounding is a special type of bias that can affect the validity of a study: it masks the true effect. A confounding variable is defined as a characteristic or variable that is associated with the problem and with a possible cause of the problem. See Howards (2018a; and 2018b). Such a characteristic or variable may either strengthen or weaken the apparent relationship between the problem and possible cause. Three conditions must exist for confounding to occur. These are:

- Confounding factor must be associated with both the risk factor of interest and the outcome;
- Confounding factor must be distributed unequally among the groups being compared;
- A confounder cannot be an intermediary step in the causal pathway from the exposure of interest to the outcome of interest.

Examples of confounders in epidemiological studies often include age, gender, and social class. In the example shown in Figure 1.8, cigarette smoking confounds the apparent relationship between coffee consumption and coronary heart disease and is thus said to be the confounding variable. The latter arises because persons who consume coffee are more likely to smoke than people who do not drink coffee, and cigarette smoking is known to be a cause of coronary heart disease (Beaglehole et al., 1993), (Figure 1.8).

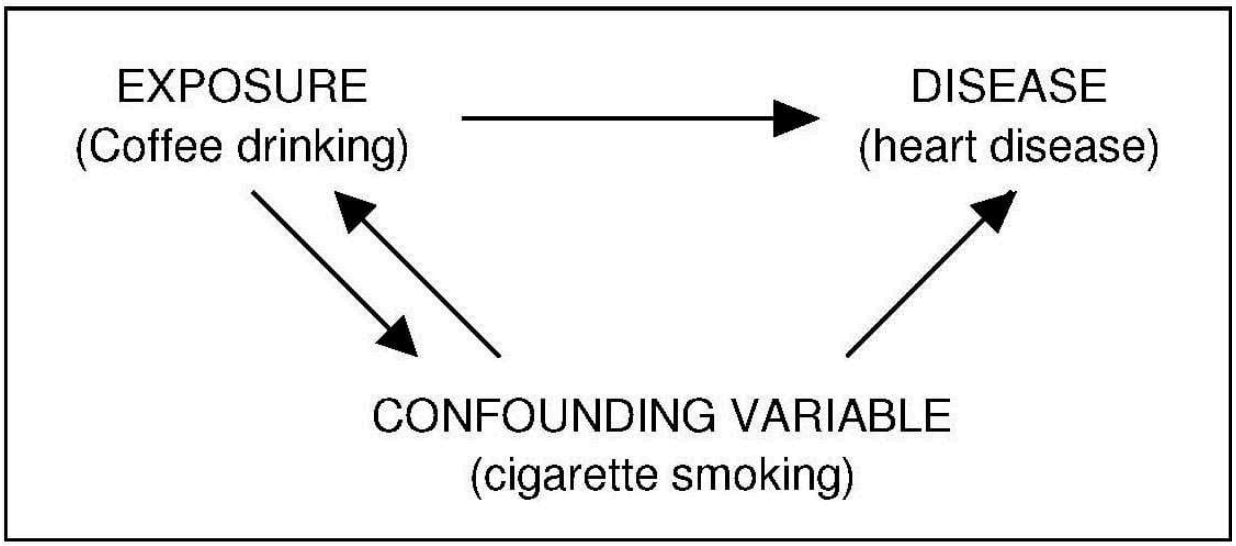


Figure 1.8: The relationship of exposure, disease, and a confounding variable. Redrawn from Beaglehole et al. (1993).

Some authors have drawn a distinction between confounders and other variables that may also influence outcome. The latter

include outcome modifiers and effect modifiers. Outcome modifiers have an effect on the health outcome independent of the exposure of interest.

Effect modifiers, in contrast, modify (positively or negatively) the effect of the hypothesized causal variables. Hence, unlike confounders and outcome modifiers, effect modifiers do lie on the causal pathway relating the exposure of interest to the outcome. As an example, hypertension is more frequent among African Americans than among Caucasians, whereas the prevalence of coronary heart disease is higher among Caucasians than among African Americans. Hence, some variable possibly related to lifestyle or constitution may modify the effect of hypertension on coronary heart disease. The number of participants needed to study effect modification is generally large, and as a consequence many studies are not powered to detect effect modification. For more details, see Newman et al. in Hulley et al. (2013) Chapter 9.

Several strategies exist to control for confounders, provided they are known and measured. They can be applied at the design or at the analysis stage, although confounding by unmeasured factors may still remain. In large studies, it is preferable to control for confounding at the analysis stage.

Strategies at the design stage include randomization to minimize the influence of baseline confounding variables, and blinding to control a biased judgement of the outcome (for RCTs only) (Section 1.6.5). Alternatively, for observational studies, restriction and matching can be used, also at the design stage; both involve changes in the sampling to ensure that only groups with similar levels of the confounders are compared.

Restriction, the simplest strategy, involves designing inclusion criteria that specify a value for the potential confounding variable, and exclude everyone with a different value. In the example depicted in Figure 1.8, if restriction was applied to avoid confounding, only nonsmokers would be included in the study design so that any association between coffee and heart disease could not be due to smoking. However, such a restriction would compromise the ability to generalize the findings to smokers, and sometimes may adversely affect recruitment, and thus the final sample size.

Matching is another strategy commonly used to control for confounding. It involves selecting individually cases and controls with the same matching values of the confounding variable(s). Both pair-wise matching and matching in groups (i.e., frequency matching) can be used. Unlike restriction, because participants at all levels of the confounder are studied, generalizability is not compromised in matching. In the Figure 1.8, when applying a case-control design, each case (i.e., person with heart disease) would be individually matched to one or more controls who smoked about the same number of cigarettes per day (i.e., pair-wise matching). The coffee drinking of each case would then be compared with that of the matched control. In some circumstances confounding variables can be controlled in the design phase without measuring them; these are termed “opportunistic observational designs”. For details, see Hulley et al. (2013)

Alternatively, when controlling for confounders at the analysis stage, potential confounders are not prespecified. Three strategies are available: stratification; statistical modeling; and propensity scores. For stratification, subjects are segregated into strata according to the level of a potential confounder, after which the relation between the predictor and outcome in each stratum are examined separately. Stratification is often limited by the size of the study, and the limited number of covariates that can be controlled simultaneously. In such cases, statistical modeling (multivariate) can be used to control multiple confounders simultaneously; a range of statistical techniques are available.

Propensity scores are used in observational studies to estimate the effect of a treatment on an outcome when selection bias due to nonrandom treatment assignment is likely. By creating a propensity score, the goal is to balance covariates between individuals who did and did not receive a treatment, making it easier to isolate the effect of a treatment; see Garrido et al. (2014). For more details of the advantages and disadvantages of strategies for coping with confounders at both the design and analysis stage, the reader is referred to Nørgaard et al. (2017). For guidelines on the appropriate use of each of these strategies, consult a statistician.

1.6.7. Sensitivity

The sensitivity of a test or indicator refers to the extent to which it reflects nutritional status or predicts changes in nutriture. Sensitive tests (or indicators) show large changes as a result of only small changes in nutritional status. As a result, they have the ability to identify and classify those persons within a population who are genuinely malnourished.

Some variables are strictly homeostatically controlled, and hence have very poor sensitivity. An example is shown in Figure 1.9. that displays the hypothetical relationship between mean plasma vitamin A and liver vitamin A concentrations. Note that plasma retinol concentrations reflect the vitamin A status only when liver vitamin A stores are severely depleted (< 0.07µmol/g liver) or

excessively high (> 1.05µmol/g liver). When liver vitamin A concentrations are within these limits, plasma retinol concentrations are homeostatically controlled and levels remain relatively constant and do not reflect total body reserves of vitamin A. Therefore, in populations from higher income countries where liver vitamin A concentrations are generally within these limits, the usefulness of plasma retinol as a sensitive biomarker of vitamin A exposure and status is limited.

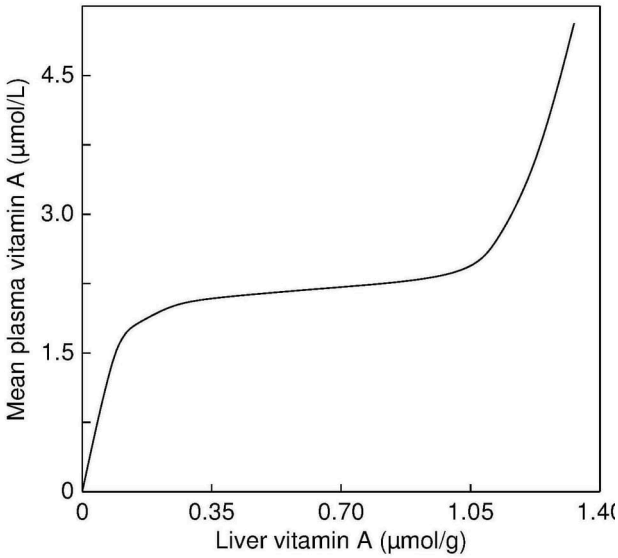


Figure 1.9: Hypothetical relationship between mean plasma vitamin A levels and liver vitamin A concentrations. Redrawn from Olson (1984).

Likewise, the use of serum zinc as a biomarker of exposure or status at the *individual* level is limited due to tight homeostatic control mechanisms. For example, doubling the intake of zinc increases plasma zinc concentrations by only 6% according to a recent meta-analysis (King, 2018).

An index (or indicator) with 100% sensitivity correctly identifies all those individuals who are genuinely malnourished: no malnourished persons are classified as "well" (i.e., there are no false negatives). Numerically, sensitivity (S_e) the proportion of individuals with malnutrition who have positive tests (true positives divided by the sum of true positives and false negatives). The sensitivity of a test (or indicator) changes with prevalence, as well as with the cutoff point, as discussed in Section 1.7.3.

Unfortunately, the term "sensitivity" is also used to describe the ability of an analytical method to detect the substance of interest. The term "analytical sensitivity" should be used in this latter context (Chapter 15).

1.6.8. Specificity

The specificity of a test (or indicator) refers to the ability of the test (or indicator) to identify and classify those persons who are genuinely well nourished. If a measurement (or indicator) has 100% specificity, all genuinely well-nourished individuals will be correctly identified: no well-nourished individuals will be classified as "ill" (i.e., there are no false positives). Numerically, specificity (S_p) is the proportion of individuals without malnutrition who have negative tests (true negatives divided by the sum of true negatives and false positives).

Table 1.6 describes the four situations that are possible when evaluating the performance of a test or indicator. These are a true positive (TP) result: the test is positive and the person really has, for example, anemia; a false-positive (FP) result: the test is positive but the person does not, for example, have anemia; a false-negative (FN) result: the test is negative but the person genuinely has anemia: and a true-negative (TN) result: the test is negative and the person does not have anemia. Increasingly in nutritional assessment, the performance of tests, and their associated indicators are being evaluated by calculating sensitivity and specificity, as well as predictive value (Section 1.6.10).

Table 1.6: Numerical definitions of sensitivity, specificity, predictive value, and prevalence for a single index used to assess malnutrit
Sensitivity (S_e) = TP / (TP+FN)
Specificity (S_p) = TN / (FP+TN)
Predictive value (V) = (TP+TN) / (TP+FP+TN+FN)
Positive predictive value (V+) = TP / (TP+FP)
Negative predictive value (V–) = TN / (TN+FN)
Prevalence (P) = (TP+FN) / (TP+FP+TN+FN)
From Habicht (1980).

Test result	The true situation: Malnutrition present	The true situation: No
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

It is important to note that sensitivity and specificity only provide information on the proportion or percentage of persons with or without malnutrition who are correctly categorized. These measures do not predict the actual number of persons who will be categorized as malnourished. The actual number of persons will depend on the frequency of malnutrition in the group being studied.

The ideal test has a low number of both false positives (high specificity) and false negatives (high sensitivity), and hence the test is able to completely separate those who genuinely are malnourished from persons who are healthy. In practice, a balance has to be struck between specificity and sensitivity, depending on the consequences of identifying false negatives and false positives. For example, for a serious condition such as screening for neonatal phenylketonuria, it might be preferable to have high sensitivity and to accept the increased cost of a high number of false positives (reduced specificity). In such circumstances, follow-up would be required to identify the true positives and true negatives.

Factors modifying sensitivity and specificity

Cutoff points have an effect on both sensitivity and specificity. In cases where lower values of the measure are associated with malnutrition (e.g., hemoglobin), decreasing the cutoff point decreases sensitivity but increases specificity for a given test. Conversely, increasing the cutoff will increase sensitivity but decrease specificity. Table 1.7 illustrates this inverse relation between sensitivity and specificity. Similarly Bozzetti et al. (1985) showed that when the cutoff for total iron binding capacity was lowered from < 310µg/dL to < 270µg/dL, the sensitivity fell from 55% to 30% but the specificity in predicting postoperative sepsis increased from 68% to 87%.

Table 1.7. Sensitivity, specificity, and relative risk of death associated with various values for mid-upper-arm circumference inchildren 6–36mos in rural Bangladesh. Data from Briend et al.(1987).

Arm circum-ference (mm)	Sensitivity (%)	Specificity (%)	Relative Risk of death
≤ 100	42	99	48
100–110	56	94	20
110–120	77	77	11
120–130	90	40	6

Extent of the random errors associated with the raw measurements influence the specificity and sensitivity of a test. If the associated random errors are large, the test will be imprecise and both the specificity and sensitivity will be reduced. Although random errors can never be completely eliminated, strategies do exist to minimize them, as noted earlier (Section 1.6.4).

Non-nutritional factors such as inflammation, diurnal variation, and the effects of disease may reduce the specificity (Habicht et al.,1979). For example, inflammation is known to decrease concentrations of serum iron, serum retinol, serum retinol binding protein, and serum zinc, while increasing serum ferritin and serum copper (Bresnahan and Tanumihardjo, 2014). As a result, the tests yield values which do not reflect true iron, vitamin A or zinc status, so misclassification occurs; individuals are designated “at risk” to low concentrations of serum iron, retinol, retinol binding protein, and serum zinc, when they are actually unaffected (false positives). In contrast, inflammation increases serum ferritin and serum copper, so that in this case individuals may be designated “not at risk” when they are truly affected by the condition (false negatives).

Table 1.8 illustrates the impact of inflammation on the geometric mean and prevalence estimates of iron, vitamin A, and zinc deficiency based on serum ferritin, retinol binding protein, and zinc.

Table 1.8. Impact of inflammation on micronutrient biomarkers of Indonesian infants of age 12mos. From Diana et al. (2017).
** Ferritin < 12µg/L*
*** RBP < 0.83µmol/L*
**** Zinc < 9.9µmol/L*

Biomarker in serum	Geometric mean (95% CI)	Proportion at risk (%)
Ferritin*: No adjustment	14.5µg/L (13.6–17.5)	44.9
Ferritin: Brinda adjustment	8.8µg/L (8.0–9.8)	64.9
Retinol binding protein**: No adjustment	0.98 (µmol/L) (0.94–1.01)	24.3
Retinol binding protein: Brinda adjustment	1.07µmol/L (1.04–1.10)	12.4
Zinc***: No adjustment	11.5µmol/L (11.2–11.7)	13.0
Zinc: Brinda adjustment	11.7µmol/L (11.4–12.0)	10.4

A new regression modeling approach has been developed to adjust serum micronutrient concentrations when affected by inflammation. In this new approach used in Table 1.8, the inflammatory biomarkers (serum C-reactive protein (CRP) and α-1-acid glycoprotein (AGP) are treated as continuous variables allowing the full range and severity of the inflammation to be accounted for; see Suchdev et al. (2016) for more details. Other disease processes may also alter the nutrient status, and in turn, the specificity of a test; for examples, see Table 15.5 in Chapter 15 (Biomarkers).

Biological and behavioral processes that relate the indicator to the outcomes may influence sensitivity and specificity. The sensitivity of low birth weight as an indicator of neonatal mortality will be greater in settings where it is due largely to prematurity rather than to intrauterine growth retardation. The sensitivity or specificity of dietary intake data collected during 24hr interviews may be affected by behavioral effects. Participants have admitted in postsurvey focus group interviews to altering their eating patterns; reasons include inconvenience, embarrassment and guilt (Macdiarmid and Blundell, 1997).

1.6.9. Prevalence

The number of persons with malnutrition or disease during a given time period is measured by the prevalence. Numerically, the actual prevalence (P) is the proportion of individuals who really are malnourished or infected with the disease in question (the sum of true positives and false negatives) divided by the sample population (the sum of true positives, false positives, true negatives, and false negatives) (Table 1.6).

Prevalence influences the predictive value of a nutritional index more than any other factor (see Section 1.6.10). For example, when the prevalence of malnutrition such as anemia decreases, it becomes less likely that an individual with a positive test (i.e., low hemoglobin) actually has anemia and more likely that the test represents a false positive. Therefore, the lower the prevalence of the condition, the more specific a test must be to be clinically useful (Hulley et al., 2013).

1.6.10. Predictive value

The predictive value can be defined as the likelihood that a test correctly predicts the presence or absence of malnutrition or disease. Numerically, the predictive value of a test is the proportion of all tests that are true (the sum of the true positives and true negatives divided by the total number of tests) (Table 1.6). Because it incorporates information on both the test and the population being tested, predictive value is a good measure of overall clinical usefulness.

The predictive value can be further subdivided into the positive predictive value and the negative predictive value, as shown in Table 1.6. The positive predictive value of a test is the proportion of positive tests that are true (the true positives divided by the sum of the true positives and false positives). The negative predictive value of a test is the proportion of negative tests that are

true (the true negatives divided by the sum of the true negatives and false negatives). In other words, positive predictive value is the probability of the person having malnutrition or a disease when the test is positive, whereas negative predictive value is the probability of the person not having malnutrition or disease when the test is negative.

The predictive value of any test is not constant but depends on the sensitivity and specificity of the test, and most importantly, on the prevalence of malnutrition or disease in the population being tested. Table 1.9 shows the influence of prevalence on the positive predictive value of an index when the sensitivity and specificity are constant. When the prevalence of malnutrition is low, even very sensitive and specific tests have a relatively low positive predictive value. Conversely, when the prevalence of malnutrition is high, tests with rather low sensitivity and specificity may have a relatively high positive predictive value (Table 1.9).

Table 1.9 Influence of disease prevalence on the predictive value of a test with sensitivity and specificity of 95%. From Dempsey and Mullen (1987).

Predictive Value	Prevalence 0.1% 1% 10% 20% 30% 40%
Positive	0.02 0.16 0.68 0.83 0.89 0.93
Negative	1.00 1.00 0.99 0.99 0.98 0.97

The predictive value is the best indicator of the usefulness of any test of nutritional status in a particular circumstance. An acceptable predictive value for any test depends on the number of false-negative and false-positive results that are considered tolerable, taking into account the prevalence of the disease or malnutrition, its severity, the cost of the test, and, where appropriate, the availability and advantages of treatment. In general, the highest predictive value is achieved when specificity is high, irrespective of sensitivity (Habicht, 1980).

Sometimes, laboratory measurements are combined with measurements of nutrient intakes and anthropometric measurements to form a multiparameter index with an enhanced predictive value. Several examples of multiparameter indices used to identify malnourished hospital patients and predict those who are at nutritional risk are discussed in detail in Chapter 27. Of these, the Nutritional Risk Index (NRI), developed by the Veterans Affairs Total Parenteral Nutrition Cooperative Study Group (1988) uses a formula that includes serum albumin level, present weight, and usual weight:

$$\begin{aligned}\text{NRI} = &(1.519 \times \text{ serum albumin }) \\ &+ 41.7 \times (\text{ present weight } / \text{ usual weight })\end{aligned}$$

The NRI was found to be sensitive and specific and a positive predictor for identifying patients at risk for complications in a study of 395 surgical patients (Veterans Affairs TPN Co-operative Study Group, 1991). NRI > 100 indicated not malnourished; NRI 97.5–100, mild malnutrition;NRI 83.5–97.5, moderate malnutrition; NRI < 83.5, severe malnutrition.

Increasingly, multiparameter indices based on three nutritional biomarkers of iron status (in combination with CRP and AGP — biomarkers of inflammation) are being used to identify iron deficiency and iron deficiency anemia at the individual and population level. The biomarkers recommended include serum ferritin, soluble transferrin receptor, and hemoglobin (Pfeiffer and Looker, 2017).

This page titled 1.6: Important characteristics of assessment measures is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

1.7: Evaluation of nutritional assessment indices

In population studies, nutritional assessment indices can be evaluated by comparison with a distribution of reference values from a healthy population (if available) using percentiles, standard deviation scores (Z-scores), and in some cases, percent-of-median (See Chapter 13 for more details). Alternatively, for classifying individuals, the values for nutritional assessment indices can be compared with either statistically predetermined reference limits drawn from the reference distribution for a healthy population or cutoff points. The latter are based on data that relate the levels of the indices to low body stores of the nutrient, impaired function, clinical signs of deficiency, morbidity or mortality. Sometimes, more than one reference limit or cutoff point is used to define degrees of malnutrition (e.g., undernutrition, overweight, obesity with body mass index) (Chapters 9 and 10).

1.7.1. Reference distribution

Reference values are obtained from the reference sample group. The distribution of these reference values form the reference distribution. The relationship between the terms used to define reference values is shown in Box 1.11

Box 1.11 The relationship between the reference population, the reference distribution, and reference limits

- REFERENCE INDIVIDUALS
  ↓ make up a
- REFERENCE population
  ↓ from which is selected a
- REFERENCE SAMPLE GROUP
  ↓ on which are determined
- REFERENCE VALUES
  ↓ on which is observed a
- REFERENCE Distribution
  ↓ from which are calculated
- REFERENCE LIMITS
  ↓ that may define
- REFERENCE INTERVALS

From IFCC (1984).

Theoretically, only healthy persons are included in the reference sample group. However, few “true” healthy reference distributions have been compiled. Exceptions include the distributions of growth reference values for the new WHO Child Growth Standards for children aged 0–60mos.These describe the growth of children whose care has followed recommended health practices and behaviors associated with healthy outcomes. Hence, they are said to be *prescriptive*, depicting physiological human growth for children 0–60mos under optimal conditions; for more details see de Onis et al. (2004). The distribution of reference values for hemoglobin (by age, sex, and race) compiled from U.S. NHANES III (1988–1991), and for serum zinc (by age, sex, blood collection time/fasting status) compiled from U.S. NHANES 1I (1976-1980), are other examples of a “true” healthy reference distributions. They are based on a sample of healthy, nonpregnant individuals, with data from any person with conditions known to affect iron status, or in the second case serum zinc concentrations, excluded (Looker et al., 1997; Hotz et al., 2003). In practice, however, more frequently the reference values for the reference sample group are drawn from the general population sampled during a nationally representative survey such as U.S. NHANES III (1988–1994) or the U.K. National Diet and Nutrition surveys.

For comparison of the observed values at the individual or population level with data derived from the reference sample, the person(s) under observation should be matched as closely as possible to the reference individuals by the factors known to influence the measurement (Ritchie and Palomaki, 2004). Frequently, these factors include age, sex, race, and physiological state, and, depending on the variable, they may also include exercise, body posture, and fasting status. In Figure 1.10, the distribution of length/height-for-age scores of male children participating in the Indian National Family Health Survey (2005–2006) are matched by age and sex with the WHO Child Growth Standard for children 0–5y. The time of day used for specimen collection

is especially critical for comparison of serum zinc concentrations with reference data (Hotz et al., 2003). Only if these matching criteria are met can the observed value be correctly interpreted. Figure 1.10

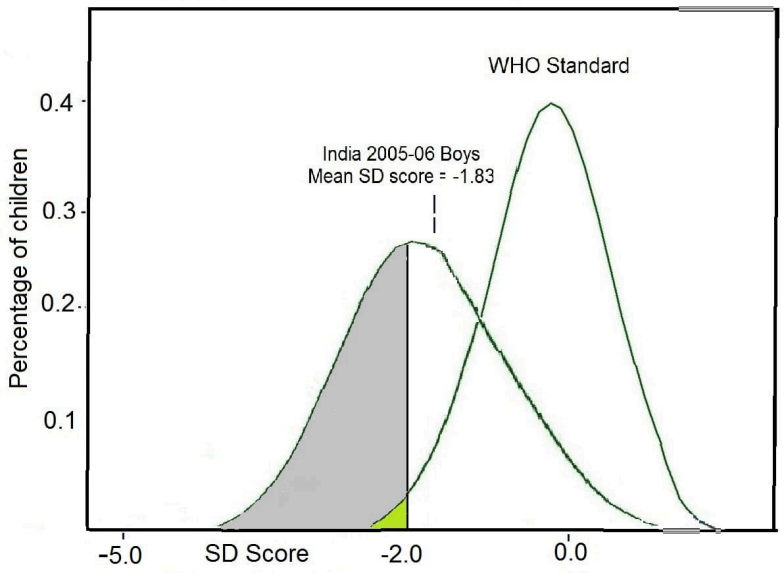


Figure 1.10 The distribution of length/height-for-age Z-scores of male children from the Indian National Family Health Survey 2005–2006. Modified from de Onis and Branca(2016)

1.7.2. Reference limits

The reference distribution can also be used to derive reference limits and a reference interval. Reference limits are generally defined so that a stated fraction of the reference values would be less than or equal to the limit, with a stated probability. Two reference limits may be defined statistically, and the interval between and including them is termed the “reference interval”. Statistically, the reference interval is often the central 95% of a normal reference distribution, and is assumed to represent a normal range. Observed values for individuals can then be classified as “unusually low”, “usual”, or “unusually high,” according to whether they are situated below the lower reference limit, between or equal to either of the reference limits, or above the upper reference limit(IFCC,1984).

In low income countries, reference limits for anthropometric growth indices based on Z-scores are preferred, with Z-scores below –2 or above +2 often used as the reference limits. to designate individuals with either unusually low or unusually high anthropometric indices. When this approach is used, theoretically the proportion of children with a Z-score less than –2 or greater than +2 in a study population should be about 2.3%. Clearly, if the proportion in the study population with such low or high Z-scores is significantly greater than this, then the study population is seriously affected, as shown in Figure 1.10. The use of Z-scores is recommended in low income countries because Z-scores can be calculated accurately beyond the limits of the original reference data. In contrast, in industrialized countries, the 3rd or 5th and 95th or 97th percentiles are frequently the reference limits used to designate individuals with unusually low or unusually high anthropometric growth indices.

Often for biochemical indices, only a lower reference limit is defined. In U.S. NHANES III, the lower reference limit for hemoglobin corresponded to the 5th percentile of the “true” healthy reference distribution of the U.S. NHANES III (1988–1991) survey, whereas for serum zinc, the lower reference limits (by age, sex, fasting status, and time of blood collection) are based on the 2.5^{th} percentile values from a “true” healthy reference sample derived from U.S. NHANES II (1976–1980). See Chapter 24 and Hotz et al. (2003).

Note that the terms “abnormal,” or “pathological” should not be applied when using this statistical approach for setting the reference limits because an unusually high or low value for an index is not necessarily associated with any impairment in health status (Smith et al., 1985).

1.7.3. Cutoff points

Cutoff points, unlike statistically defined reference limits, are based on the relationship between nutritional assessment indices and low body stores, functional impairment, or clinical signs of deficiency or excess, as noted earlier (Raghaven et al. 2016). Their use is less frequent than that of reference limits because information relating nutritional assessment indices and functional impairment or clinical signs of deficiency or excess is often not available. Figure 1.11 depicts the global prevalence of overweight and obesity for adult males and females based on a population measure (i.e., cutoff for BMI ≥ 25 by age and sex). In this example, a BMI cutoff defined as 25 or higher is based on the evidence that excess weight is associated with an increased incidence of cardiovascular diseases, type 2 diabetes mellitus, hypertension, stroke, dyslipidemia, osteoarthritis, and some cancers (Burton et al., 1985).Cutoff points may vary with the local setting because the relationship between the nutritional indices and functional outcomes is unlikely to be the same from area to area.

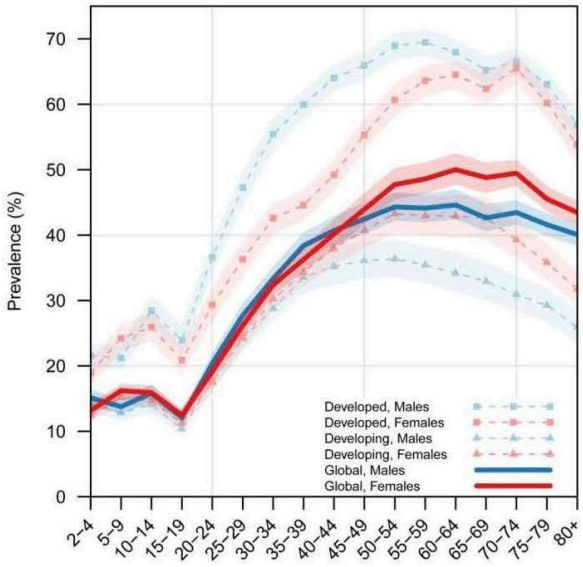


Figure 1.11: Prevalence of overweight and obesity (BMI ≥ 25) by age and sex, 2013. Modified from: Ng et al. (2014).

Cutoff points, like reference limits, are often age-, race-, or sex-specific, depending on the index. They must also take into account the precision of the measurement. Poor precision affects the sensitivity (Section 1.6.7) and specificity (Section 1.6.8) of the measurement, and leads to an overlap between those individuals classified as having low or deficient values with those having normal values. This results in misclassification of individuals.

Sometimes more than one cutoff point is selected. For example, several cutoffs based on body mass index (BMI, kg/m^2) are used to classify the severity of overnutrition in adults (see Chapter 10), whereas for serum vitamin B_{12} two cutoffs associated with vitamin B_{12} deficiency or depletion have been defined (Allen et al., 2018). The U.S. Institute of Medicine have published four cutoffs for serum total 25-hydroxyvitamin D to define four stages of vitamin D status: deficiency; insufficiency; sufficiency; no added benefit; possible harm (see Chapter 18b for more details), with the limit for deficiency (< 12ng/mL; < 30nmol/L) based on relationships to biomarkers of bone health (Ross et al., 2011).

When selecting an index and its associated cutoff point, the relative importance of the sensitivity and specificity of the nutritional index (or indicator) must always be considered, as noted earlier (Sections 1.6.7, 1.6.8). Receiver operator characteristic (ROC) curves are often used to select cutoff points. This is a graphical method of comparing indices and portraying the trade-offs that occur in the sensitivity and specificity of a test when the cutoffs are altered. To use this approach, a spectrum of cutoffs or thresholds over the observed range of results is required and the sensitivity and specificity for each cutoff calculated. Next, the sensitivity (or true-positive rate) is plotted on the vertical axis against the true negative rate (1.0–specificity) on the horizontal axis for each of the three cutoff points as shown in Figure 1.12. The closer the curve follows the left-hand border and then the top-border of the ROC space, the more accurate is the biomarker cutoff in distinguishing a deficiency from optimal status. The optimal ROC curve is the line connecting the points highest and farthest to the left of the upper corner. The closer the curve comes to the 45° diagonal of the ROC space, the less accurate the biomarker cutoff. Most statistical programs (e.g., SPSS) provide some sort of ROC curve analysis (Søreide, 2009). For more details, see Chapter 15.

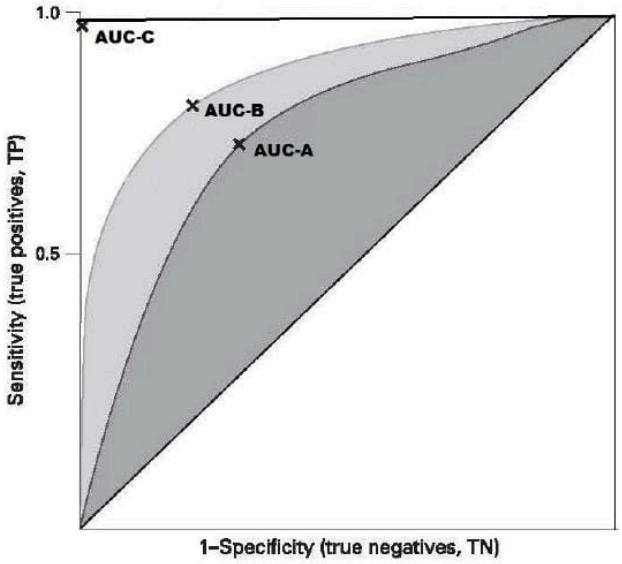


Figure 1.12. Receiver-operating characteristic curves. Three plots and their respective areas under the curve (AUC) are given. The diagnostic accuracy of marker C (white area) is better than that of B and A, as the AUC of C > B > A. X = optimal cutoff point for each of the three markers. Redrawn from: Søreide (2009).

The area under the ROC curve (AUC), also known as the cut-point "c" statistic or c index, is a commonly used summary measure of the accuracy of the cutoff for the nutritional assessment index of interest. AUCs can range from 0.5 (random chance, or no predictive ability); this follows the 45° line in the ROC plot, see (Figure 1.10) to > 0.75 (good), and > 0.9 (excellent). The cutoff value that provides the highest sensitivity and specificity is calculated. On the rare occasions that the estimated AUC for the index cutoff is < 0.5, then the index cutoff is worse than chance. When multiple indices are available for the same nutrient, the index with the highest AUC is often selected.

The Youden index (J) is another main summary statistic of the ROC curve. It defines the maximum potential effectiveness of a biomarker. J can be defined as:

$$\mathrm{J} = (\ \text{maximum sensitivity}\ (\mathrm{c}) + \ \text{specificity}\ (\mathrm{c}) - 1$$

The cutoff that achieves the maximum is referred to as the optimal cutoff (c*) because it is the cutoff that optimizes the biomarker's differentiating ability when equal weight is given to sensitivity and specificity. J can range from 0 to 1, with values closer to 1 indicating a perfect diagnostic test and values closer to 0 signifying a limited effectiveness. For more details, see Schisterman et al. (2005) and Ruopp et al. (2008).

Misclassification arises when there is overlap between individuals who actually have the deficiency (or excess) and those falsely identified (i.e., false positives). Neither reference limits nor cutoff values can separate the "deficient" and the "adequately nourished" without some misclassification occurring. This is shown in Figure 1.13 for the real-life situation (B).

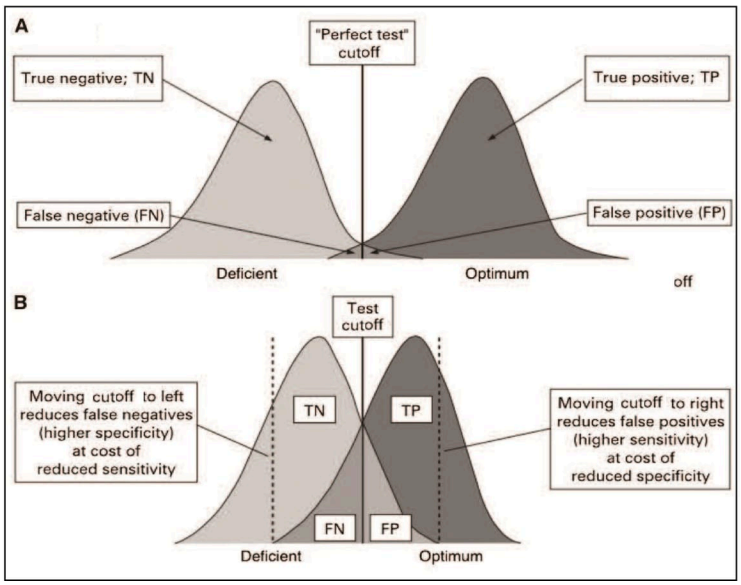


Figure 1.13. A good discriminatory test with almost perfect ability to discriminate between people with a nutrient deficiency and those with optimum nutrient status. The ability to correctly detect all the true negatives depends on the specificity of the biomarker; the ability to correctly detect all the true positives depends on the sensitivity of the biomarker. FN, false negative; FP, false positive; TN, true negative; TP, true positive. Redrawn from: Raghaven et al. (2016).

Note that the final selection of the cutoff values may vary depending on whether the consequences of a high number of individuals being classified as false positive is more or less important than the consequences of a large number of individuals being classified as false negatives. Minimizing either misclassification may be considered more important than minimizing the total number of individuals misclassified.

Note that the sensitivity can be improved (i.e., reducing the false positives) by moving the cutoff to the right, but this reduces the specificity (false negatives), whereas moving the cutoff to the left reduces the false negatives (higher specificity) at the cost of a reduction in sensitivity. The former scenario may be preferred for the clinical diagnosis of a fatal condition, whereas cutoffs with a high specificity may be preferred for diagnostic tests that are invasive or expensive.

Misclassification arises because there is always biological variation among individuals (and hence in the physiological normal levels defined by the index), depending on their nutrient requirements(Beaton, 1986). As well, for many biomarkers there is high within-person variation, which influences both the sensitivity and specificity of the index, as well as the population prevalence estimates. These estimates can be more accurately determined if the effect of within-person variation is taken into account. This can only be done by obtaining repeated measurements of the index for each individual, which for invasive biochemical biomarkers, is often not feasible.

Figure 1.13 illustrates the problem of misclassification. In this figure, the light-shaded area to the right of 110g/L and below the left curve represents anemic persons classified as normal according to the cutoff point (110g/L) defined by the World Health Organization (WHO, 1972). The dark-shaded area to the left of 110g/L and below the right curve comprises persons within the normal population, classified as anemic by the WHO cutoff point but who were not found to be responsive to iron administration. Hence, the dark-shaded area represents those well-nourished persons who were incorrectly classified as “anemic” (i.e., false positives).

1.7.4. Trigger levels for surveillance and public health decision making

In population studies, cutoff points may be combined with trigger levels to set the level of an index (or indicator) or combination of indices at which a public health problem exists of a specified level of concern. Trigger levels may highlight regions or populations, where specific nutrient deficiencies are likely to occur, or may serve to monitor and evaluate intervention programs. They should, however, be interpreted with caution because they have not necessarily been validated in population-based surveys.

Some international organizations including WHO and UNICEF (2018), the International Vitamin A Consultative Group (Sommer and Davidson, 2002),and the International Zinc Nutrition Consultative Group (IZiNCG, 2004), for example, have defined the prevalence criteria for selected indicators within a population that signify a public health problem in relation to specific nutrients

and conditions.

As an example, the WHO and UNICEF have classified the severity of malnutrition in young children age < 60mos into five thresholds based on the prevalence (as %) of wasting (i.e., weight-for-length/height < –2 Z-scores), overweight (weight-for-age > +2 Z-scores), and stunting (length/height-for-age < –2 Z-scores) for targeting purposes (de Onis et al., 2018). The number of countries (i.e. "n") in each of the different threshold categories, based on data from the WHO 2018 Global Database on Child Growth and Malnutrition, are also shown (Table 1.10).

Table 1.10. Prevalence thresholds, corresponding labels, and the number of countries (n) in different prevalence threshold categories for wasting, overweight and stunting in children under 5 years using the "novel approach". From de Onis et al. (2018).

Wasting			overweight			Stunting		
Prevalence thresholds (%)	Labels	(n)	Prevalence thresholds (%)	Labels	(n)	Prevalence thresholds (%)	Labels	(n)
< 2·5	Very low	36	< 2·5	Very low	18	< 2·5	Very low	4
2·5 – < 5	Low	33	2·5 – < 5	Low	33	2·5 – < 10	Low	26
5 – < 10	Medium	39	5 – < 10	Medium	50	10 – < 20	Medium	30
10 – < 15	High	14	10 – < 15	High	18	20 – < 30	High	30
≥ 15	Very high	10	≥ 15	Very high	9	≥ 30	Very high	44

Box 1.12. Trigger levels for zinc biomarkers

- Prevalence of serum zinc less than age/sex/time-of-day specific cutoffs is > 20%
- Prevalence of inadequate zinc intakes below the appropriate Estimated Average Requirement (EAR) is > 25%
- Prevalence of low height-for-age or length-for-age Z-scores (i.e., < –2SD) is at least 20%.

Comparison of the prevalence estimates for each anthropometric indicator can trigger countries to identify the most appropriate intervention program to achieve "low" or "very low" prevalence threshold levels. Trigger levels for zinc biomarkers have been set by the International Zinc Nutrition Consultative Group. (Box 1.12) Note that ideally, all three types of indicators should be used together to obtain the best estimate of the risk of zinc deficiency in a population and to identify specific sub-groups with elevated risk (de Benoist et al., 2007).

A generalized discussion of the specific procedures used for the evaluation of dietary, anthropometric, laboratory, and clinical methods of nutritional assessment are discussed more fully in Chapters 8, 13, 15, and 25, respectively.

This page titled 1.7: Evaluation of nutritional assessment indices is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

2: FOOD AVAILABLE FOR CONSUMPTION AT NATIONAL AND HOUSEHOLD LEVELS

2.1: Introduction
2.2: Food Balance Sheets
2.3: Household Surveys
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex

This page titled 2: Food available for consumption at National and Household Levels is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Introduction

This chapter considers survey methods suitable for a quantitative assessment of food available for consumption / consumed at national and household levels. Methods suitable for measuring food consumption at the individual level are discussed in Chapter 3. Methods designed to estimate the exposure to hazardous substances, although sometimes based on household food supply, and those providing only a qualitative assessment such as the household dietary diversity score, or capturing only part of the available food supply (e.g., using Universal Product Codes — UPCs), are not considered.

The food available for human consumption by a population can be assessed through nationally representative surveys at the national level using Food balance Sheets (FBS), or alternatively at the household level using either Household Food Consumption Surveys (HFCS) or Household Consumption and Expenditure Surveys (HCES). The resulting data from all these surveys are expressed as per capita food supply for human consumption, and are used to estimate per capita available intake of foods, and energy and nutrients, when matched with appropriate food composition data.

The term that should be used for data derived from most of these national and household surveys is either “food supply” or “food available for consumption” rather than “food consumption”. Only certain types of household surveys measure *actual* foods consumed.

Data from FBSs provide no information on the distribution of the food available for consumption within the country (in terms of geographical areas, seasonality, or demographic characteristics), unlike the data generated from household surveys. However, household surveys (HCS or HCES) do not provide information on the availability of food/nutrients by age, sex, and physiological status. Moreover, the interpretation of the data from household surveys in terms of dietary adequacy presents challenges and uncertainties unless complemented with individual-level dietary intake data.

Several indicators can be derived from FBS and HCES data to monitor and compare food insecurity at the global and national levels. These indicators can be based on the quantity of food available for consumption (e.g., prevalence of undernourishment (PoU) based on the energy supply per capita), as well as the food quality (e.g., Household Dietary Diversity Score).

This page titled 2.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Food Balance Sheets

Food balance sheets are particularly important in countries where food consumption surveys at the household or individual level are not performed regularly on a representative sample of the population, as in some low-income countries.

Several terms have been used to describe FBSs. These include “country's food supply”, “national food accounts”, “food moving into consumption”, “food consumption statistics”, “food disappearance data”, “food available for consumption” and “consumption level estimates”. Here, the term “country's food supply” is used.

Food balance sheets present a comprehensive picture of the pattern of a country's food supply during a specified reference period: the calendar year, the agricultural year or the crop year (FAO, 2021c). The sources of the food supply and its utilization are reported for each food item potentially available for human consumption, and includes primary commodities such as ‘wheat, rice, fruit, vegetables, and processed commodities such as vegetable oils and butter. FBS data cannot be disaggregated to determine the distribution of a country's food supply spatially, seasonally, or by demographic characteristics.

Both official and unofficial data are used for the compilation of FBSs; National statistical offices constitute an important data source. However, data measured directly on food availability at national level may be difficult to obtain (FAO, 2021a). Instead, FBS compilers often derive estimates of food availability by making certain adjustments based on other existing data sets that measure food production or consumption. For example, two official data sources that may assist in the estimation of a country's food supply availability are Industrial Output Surveys and Household Consumption and Expenditure Surveys (HCES). See Section 2.2 for the description of HCES surveys.

The country's food supply is calculated from national food production, plus imports, plus food taken from stocks. Exports and food added to stocks are then subtracted to obtain a gross estimate of a country's food supply. The estimates of food diverted for non-human uses (e.g. animal feed, seed, non-food use) and of food loss are then subtracted from the gross country's food supply to obtain the net country's food supply, i.e. the total food available for human consumption in a country at the retail level (Figure 2.1, Nelson, 1984). Data on non-commercial food production and detailed information on processed foods are not available in Food Balance Sheets.

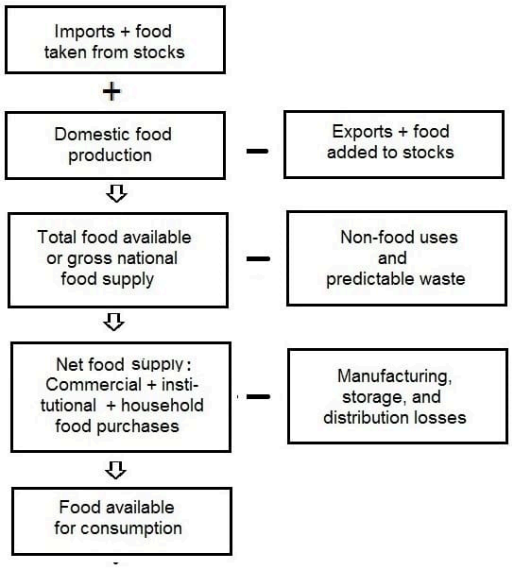


Figure 2.1. The derivation of food balance sheets. Losses at the retailer, catering and household level and non-commercial food production are not taken into consideration when compiling food balance sheets. Modified from Nelson (1984).

For the purposes of the FBS, “food loss” closely aligns with “post-harvest / post-slaughter loss”. It represents the quantity of food that leaves the production / supply chain at any stage, e.g., during production, processing and distribution, up to the retail level (FAO, 2021a). Food loss may be considerable in countries where the agricultural products reach the consumer after travelling long distances and/or passing through several intermediaries prior to retail. In such cases, large amounts of food are lost because of damage, especially perishable foods such as fresh fruit and vegetables or processed foods with a short shelf life. Assumptions

about losses may be based on expert opinion obtained in the countries. Losses or waste at the retailer, food catering, and household level are not considered in the compilation of FBS. Hence, total food supply estimates derived from the FBS are likely to be higher than average food consumed at the household level.

An example of the upper portion of an FAO food balance sheet which also shows in detail the sources of domestic supply and utilization of two of the major food groups (i.e., Cereals and Starchy roots) is shown in Table 2.1.

Item	Pop.	Domestic Supply					Domestic Utilisation							Per Capita Supply			
		Prod.	Imp.	Stock Var.	Exp.	Total	Food	Proc.	Feed	Seed	Losses	Oth. Use	Tourist	Resid.	Total		Prot.	Fat
	(1000)	(1000 tonnes)													Kg/Yr	KCal/Day	g/Day	
Population	81,116																	
Grand Total																3,540	101	126
Vegetal Products																3,002	65	92
Animal Products																538	36	34
Cereals - Excl. Beer		**36,178**	**8,590**	**2,629**	**7,298**	**34,840**	**14,791**	**510**	**12,899**	**2,000**	**2,621**	**2,018**		**0**	**182**	**1,366**	**42**	**5**
Wheat and products		21,500	5,178	2,599	6,449	17,630	11,655	20	2,093	1,380	1,205	1,276		0	144	1,086	36	4
Rice and products		900	448	-88	93	1,343	1,294		1	14	33	1		0	16	98	2	0
Barley and products		7,100	460	147	9	7,404		174	5,608	520	1,065	38		0				
Maize and products		5,900	2,492	-39	694	7,737	1,643	221	4,898	47	224	704		0	20	162	4	1
Rye and products		320		15	0	305	198		39	20	48			0	2	20	1	0
Oats		250	1	0	0	251	1	95	103	15	37			0	0	0	0	0
Millet and products		5	7	0	0	11			11	0	0			0				
Sorghum and products		0		0	0	0					0			0				
Cereals, Other		203	4	-5	54	159	0		146	4	9			0	0	0	0	0
Starchy Roots		**4,801**	**305**	**1**	**285**	**4,820**	**3,801**	**0**	**259**	**264**	**486**	**10**		**0**	**47**	**91**	**2**	**0**
Cassava and products			259		0	259	0		259			0		0	0	0	0	0
Potatoes and products		4,800	46	1	285	4,560	3,800	0	0	264	486	10		0	47	91	2	0
Sweet potatoes		0	0	0		0			0		0			0				
Roots, Other		1	0	0	0	1	1		0		0			0	0	0	0	0

Table 2.1. FBS - FAOSTAT.
Abbreviations: Pop, Population; Prod, Production, Imp, Imports; Exp, Exports; Proc, Processed commodities converted back to their original commodity equivalent; Oth. Use, Other Uses; Tour, Tourists; Resid, Residual. From FAO, (2022a).

In some cases, information on a country's food stocks is not available. The effect of this data gap is reduced by preparing FBSs as a moving average for 3y periods for all countries. For more details, see FAO (2021f).

FAO in its online database (FAOSTAT) provides free access to FBSs for over 245 countries and territories and all the FAO regional groups, from 1961 to date (FAO, 2021c). The FBSs reported are “standardized FBSs” in which the commodity list is confined only to the quantity of the primary commodities available for consumption, with the quantity of processed commodities converted into their original primary commodity equivalent. For example, for bread, quantities are expressed in wheat equivalents. For more details on the standardization of Food Balance Sheets, see FAO (2021g).

The major food groups listed in the FAO FBS data are shown in Box 2.1. A list of food commodities classified into these major food groups is available from FAO (2021c) in the definition and standard section of the FBS website. As an example, the food commodity “cassava” can be found in the food group “starchy roots”.

Box 2.1. Major food groups listed in the FAO Food Balance Sheet data

- Cereals - Excluding Beer
- Starchy Roots
- Sugar Crops
- Sugar & Sweeteners
- Pulses
- Treenuts
- Oilcrops
- Vegetable Oils
- Vegetables
- Fruits - Excluding Wine
- Stimulants
- Spices

- Alcoholic Beverages
- Meat
- Offals
- Animal fats
- Eggs
- Milk - Excluding Butter
- Fish, Seafood
- Aquatic Products, Other
- Miscellaneous

From FAO, 2021c.

The per capita supply available for human consumption is obtained by dividing the food supply by the estimate of the partaking population in each country; it is expressed in grams per capita of individual food commodities. FAO uses mid- year official population estimates released by the United Nations Development Program. FAO recommendation for FBS compilers, the number of migrants and tourists must be calculated as the difference between country's outbound travellers and inbound visitors and then subtracted from the population estimate (FAO, 2021a).

Data on per capita food supplies are expressed in terms of quantity, and in terms of caloric value, referred to as the Dietary Energy Supply (DES) (kcal/capita/day). The protein and fat content of each food group and commodity calculated by applying appropriate food composition data are also available and expressed in terms of g/capita/day (Table 2.1). In the future, FAO may also provide micronutrient data for each food group and commodity.

Food balance sheets are derived statistics. Hence, their accuracy is dependent on the reliability of the underlying basic statistics. The coverage and accuracy of underlying statistics vary markedly across countries. In some low-income countries, the coverage and quality of the statistics are uncertain — especially for food diverted for non-human food uses and for food stocks. As well, in low-income countries where subsistence agriculture is widespread, FBSs may result in an underestimate of per capita food supply because the household consumption of vegetal food products obtained through growing or gathering, and animal food products through breeding or hunting, are not considered in the FBS calculations. Further, as food systems become more sophisticated in countries, systematic errors may increase.

In FAO documents, it is made clear that availability for human consumption does not equate with consumption. The quantities of food available for human consumption, as estimated in the FBS, reflect only the quantities reaching the consumer at retail level. The amount of food (and of nutrients) actually consumed may be lower than the quantity shown depending on the degree of loss and waste of edible food and nutrients in the household. These may occur during storage, in preparation and cooking, as plate-waste, and when fed to domestic animals and pets, or when thrown away. Storage, preparation, and cooking will have a greater effect on the content of vitamins and minerals of the food commodities than on the protein and fat content. See: FAO (2021h).

Recently, FAO has changed its FBS methodology, details of which are described in the document “The New Food Balances and the utilization variables” (FAO, 2021a). The key differences between the new and old FBS include a proportional balancing mechanism, use of the 2019 UNPD population data, inclusion of a new food module, revised computations for losses, feed, and stocks, and addition of a new element on non-food for industrial use; for more details see (Lancet, 2024).

The release of FBS data for 2019 includes this new methodology for the FBS data for 2014–2019. The new methodology has also been applied backwards to cover 2010–2013, thereby giving consistency over time for the national per capita food supply data from 2010 to 2019. In the new methodology, food available for consumption by non-resident visitors has also been included for selected countries in which tourism has a significant impact on the food supply (e.g., Small Island Developing States). Food Balance Sheet data preceding 2010 will soon be recompiled using the new methodology. FAO has developed a FBS capacity development package for countries that includes a compilation tool incorporating the new methodology. Details are given in (FAO, 2022a).

2.2.1 Uses of Food Balance Sheet data

Food Balance Sheet data are easily accessible and provide coverage of per capita food supply across countries and regions using a relatively consistent methodology compared to many other sources of data on food supply. Hence, they are often used to analyze overall trends, historic trends, and changes at country, regional, and global levels, in relation to both Dietary Energy Supply (Figure 2.2) and Food Groups (Figure 2.3). Asia has experienced the fastest growth in DES since 2010 (i.e., 6%), whereas

in contrast the lowest historical DES is apparent in Africa, where there has been a slightly decreasing trend over the last decade (Figure 2.2).

Food Balance Sheets can also be used to formulate agricultural policies concerned with the production, distribution, and consumption of foods.

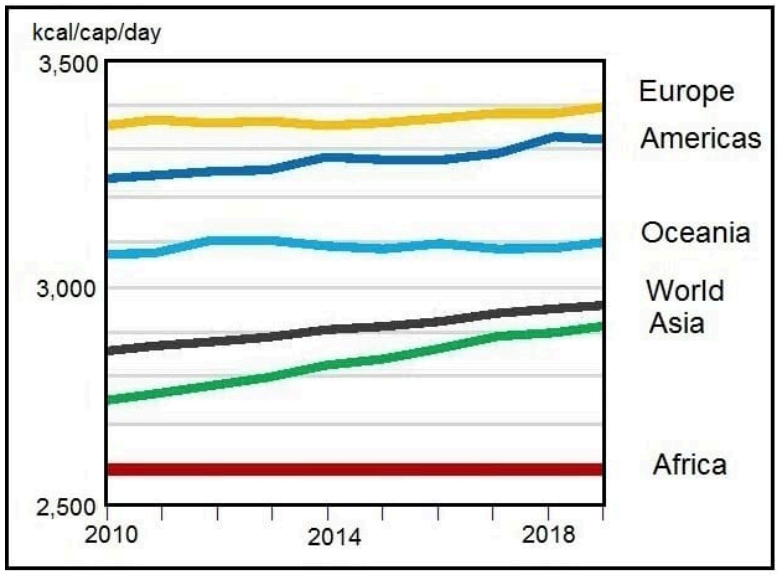


Figure 2.2 Dietary Energy Supply (DES) by region from 2010 to 2019 (kcal/cap/day). From FAO (2021d).

In relation to food groups, at the global level in 2019, cereals accounted for the highest share of the total food supply, both in quantities (in tonnes) available (25%) and dietary energy (44%), whereas although fruit and vegetables represented 31% of the total quantities available, they only accounted for 7% of the total dietary energy available due to their low kilocalorie content. In contrast, sugar, and fats and oils represented only 4% and 3% of the total quantities available but 8% and 14% respectively of the dietary energy available, as shown in Figure 2.3.

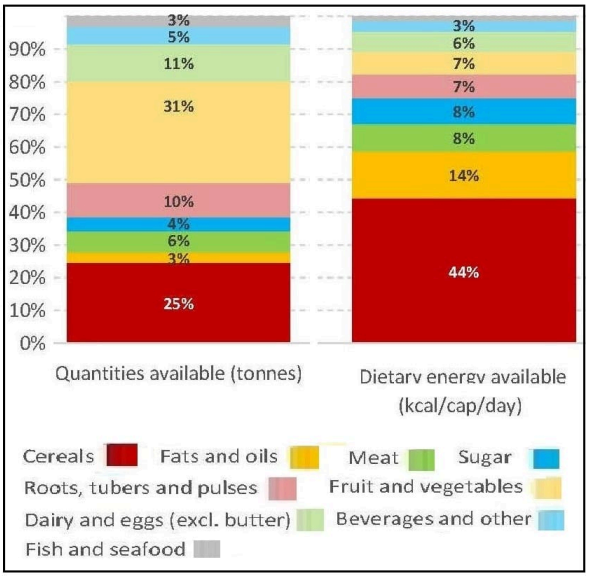


Figure 2.3. Global food availability composition (2019) From FAO (2021d).

In addition, FBS data have many other uses, including correlations with other data which have a global coverage. Thar et al. (2020) have summarized several health-related uses of FBS data, as shown in Box 2.2, based on their review of 119 FBS surveys. They identified 50 studies conducted from the earliest available until 2016 that examined associations between dietary factors and mortality or health outcomes such as cancer, diabetes, or obesity. However, any interpretation in terms of causal/effect of such observed associations is highly questionable and should be discouraged. Further, given the importance of the decisions

that may be taken based on FBS data, users need to have a good understanding of the strengths and weaknesses of the source of the data. See Box 2.2 and Table 2.7 and the review of Thar et al. (2020) for a summary of some of the appropriate uses and potential weaknesses of FBS data.

Box 2.2. Health-related uses if food balance sheet data

- Inter-country, regional or global comparison of food groups, nutrient availability and/or dietary trends
- Within-country analysis of food availability, nutrient availability and/or dietary trends
- Comparison of FBS with other national sources of dietary data
- Modeling studies in which FBS data are used as a guide to assess whether a proposed intervention is likely to make a difference to the health and nutrition of the population. As an example, Sheehy and Sharma (2011). investigated the possibility of increasing the folate supply in the European Union by replacing normal eggs with folate-enriched eggs
- Association between FBS dietary factors and mortality or health outcomes.

From Thar et al (2020).

Several investigators have examined the validity of the global dietary estimates generated from FAO FBS data. For example, Del Gobbo et al. (2015)compared the FAO data with estimates based on nationally representative dietary data from individual-level surveys from 113 countries. Substantial overestimates were reported for most food groups including fruits, vegetables, whole grains, red and processed meat, fish and seafood, milk, and total energy, whereas beans and legumes, nuts and seeds were under-estimated based on FAO national food-supply data compared to the corresponding estimates from the national data at the individual-level. As a result, calibration equations were developed to adjust the FAO estimates to improve their validity. The authors suggest that these calibration models could be used not only to improve estimates based on FBS data of per capita food intakes at the national level but also by age and sex.

One of the main uses of FBS is to monitor global hunger and food security by providing data on national diets and their nutrient adequacy. With these, insights into the countries and regions that are most likely to be at greatest risk for low dietary supply of energy (DES)(kcal per capita) as well as selected nutrients, can be provided. Selected indicators derived from the food component data of FBS and used to monitor global hunger, food security, and nutrient adequacy are presented in Box 2.3. Details of some of their applications are given below.

Box 2.3. Selected global hunger, food security and nutrition indicators based on Food Balance Sheet Data

- Dietary Energy Supply (DES) (kcal/capita/day)*
- Average Dietary Energy Requirement*
- Prevalence of Undernourishment (PoU)*
- Dietary energy available from cereals, roots, and tubers (kcal/capita/day), expressed as percent*
- Protein supply (g/capita/day), 3-year average*
- Fat supply (g/capita/day), 3-year average*
- Supply of protein of animal origin (g/capita/day), 3-year average*
- Estimated prevalence of inadequate intakes of selected micronutrients at global and country-specific levels

*These indicators are present in the FAO “Suite of indicators of food security”

2.2.1a Trends in global hunger

Ideally, the assessment of global hunger (and food security) should be based on national data on food consumption at the household (see Section 2.2) or the individual level (see Chapter 3). However, only a few countries conduct such surveys on an annual basis. Therefore, FBSs have a key role in the assessment of global hunger and food insecurity, published yearly in the report “The State of Food Security and Nutrition in the World” (SOFI) by FAO, IFAD, UNICEF, WFP and WHO (2022). This global monitoring report identifies countries in which food insecurity is most prevalent, monitors global hunger and food security trends over time, and provides projections of future global hunger and food insecurity.

In this report, the per capita Dietary Energy Supply (DES) (Box 2.3 and Table 2.1) estimated from FBSs is used as a proxy for the

average energy intake of the population. The per capita Dietary Energy Supply (DES) is available from FAOSTAT and calculated using an FAO global food composition database. The sum of the dietary energy content of the total food supply is then divided by the population size of the country and by 365 to calculate the per capita daily dietary energy supply (DES) available for human consumption (See Table 2.1 and Figure 2.2). However, DES does not yield any information on the affordability, access, or consumption of dietary energy by different population groups within a country, although the contribution of energy from individual food groups (Table 2.1) available for human consumption at the global level and within a country are calculated.

Dietary Energy Supply is one of the three metrics used to estimate the prevalence of undernourishment (PoU), an indicator based on the percentage of individuals in the population who are in a condition of "undernourishment" (Table 2.2). PoU is one of the indicators of Sustainable Development Goal 2 (FAO, 2021e) and is defined by FAO as the condition of an individual whose habitual food consumption is insufficient to provide, on average, the amount of dietary energy required to maintain a normal, active, and healthy life. The national PoU estimates are reported as three-year moving averages, whereas the regional and global aggregates are reported as annual estimates. Table 2.2 highlights the regional disparities in PoU, with Africa having the highest burden of undernourishment as well as the region with the largest increase in the proportion of the population affected by undernourishment during 2005 to 2021. Disparities at the subregional levels also exist and are summarized in FAO, IFAD, UNICEF, WFP and WHO (2022). PoU, however, does not provide information on which specific individuals are undernourished or on the quality or diversity of the diet. The three metrics used to estimate PoU are:

- Average dietary energy consumption (DEC). DEC is based on DES minus the incidence of caloric losses at retail distribution level
- Coefficient of Variation (CV) of energy intake. This is a measure of the level of inequalities in access to dietary energy consumption within a population
- Average minimum dietary energy requirement (MDER) per person for the national population in each of the covered countries. This represents the minimum amount of dietary energy per person that is considered adequate to ensure they maintain a minimum weight for health.

To estimate the CV of dietary energy intake, data on inter-household variability from household surveys such as Household Consumption and Expenditure Surveys (HCES) are used. Higher CV values represent larger levels of dietary inequality. Use of such data in the PoU calculation is crucial because food (and consequently energy) is never equally distributed among households within a country. Hence, any comparison of per capita DES with energy requirements at the country level without considering the CV of energy intake would largely underestimate the prevalence of food insecurity at household level in many countries.

To derive the MDER, factors such as the age and sex structure of the population as well as height of the individuals within each country are applied. A coefficient is assumed for physical activity corresponding to the average values of the PAL category "Sedentary or light activity lifestyle" (FAO human energy requirement 2001). Using the variation in intake and energy requirements, undernourishment can be calculated. More details of how these three metrics are used to estimate PoU are given in FAO (2022).

	Prevalence of undernourishment (percent)								
	2005	2010	2015	2016	2017	2018	2019	2020*	2021*
WORLD	12.3	8.6	8.0	7.8	7.6	7.7	8.0	9.3	9.8
AFRICA	20.7	16.5	15.8	16.3	16.4	17.0	17.4	19.6	20.2
Northern Africa	8.4	6.4	5.2	5.4	5.6	5.5	5.4	5.9	6.9
Sub-Saharan Africa	23.9	18.9	18.3	18.9	18.8	19.6	20.1	22.7	23.2
Eastern Africa	33.8	26.5	24.4	25.2	25.4	26.6	27.5	30.2	29.8
Middle Africa	34.9	26.0	26.3	27.4	26.6	27.3	28.1	30.4	32.8
Southern Africa	4.9	5.8	7.4	7.4	7.5	7.4	7.9	9.1	9.2
Western Africa	12.2	9.9	10.1	10.1	10.0	10.6	10.4	13.2	13.9
ASIA	13.9	9.1	8.0	7.5	7.1	7.1	7.4	8.6	9.1
Central Asia	14.0	6.0	3.8	3.5	3.2	2.9	2.6	3.1	3.1
Eastern Asia	6.8	<2.5	<2.5	<2.5	<2.5	<2.5	<2.5	<2.5	<2.5
South-eastern Asia	17.2	10.9	7.8	6.7	6.0	5.9	5.6	5.8	6.3
Southern Asia	20.5	15.3	14.1	13.1	12.4	12.3	13.2	15.9	16.9
Western Asia	7.8	5.9	9.6	10.4	10.2	10.3	10.0	10.1	10.0
Western Asia and Northern Africa	*8.1*	*6.1*	*7.6*	*8.1*	*8.1*	*8.1*	*7.9*	*8.2*	*8.6*
LATIN AMERICA AND THE CARIBBEAN	9.3	6.6	5.8	6.7	6.4	6.6	6.7	8.0	8.6
Caribbean	18.7	15.2	14.2	14.5	14.4	15.2	15.2	16.5	16.4
Latin America	8.6	6.0	5.1	6.2	5.8	6.0	6.1	7.4	8.0
Central America	8.0	7.3	7.5	8.1	7.9	7.9	7.6	8.0	8.4
South America	8.8	5.5	4.2	5.4	5.0	5.2	5.4	7.1	7.9
OCEANIA	6.8	6.2	5.7	5.8	5.8	5.7	5.6	5.4	5.8
NORTHERN AMERICA AND EUROPE	<2.5	<2.5	<2.5	<2.5	<2.5	<2.5	<2.5	<2.5	<2.5

NOTES: * Projected values based on the middle of the projected range

Table 2.2. Selected data on the prevalence of undernourishment (PoU) (as percent) from 2005-2021 for the world, Africa, Asia, Latin America and the Caribbean, Oceania, and Northern America and Europe. From FAO, IFAD, UNICEF, WFP and WHO (2022).

Note the presence of persistent regional disparities in the proportion of the population affected by hunger from 2005 to 2021 (Table 2.2), with Africa having the highest burden of undernourishment as well as the region with the largest increase in the proportion of the population affected by hunger during this time. Disparities also exist at the subregional levels: these are also summarized in the report.

FAO also provides data on the Cost and Affordability of a Healthy Diet (CAHD). This has been shown to be strongly associated with PoU and with different forms of malnutrition, including child stunting and adult obesity. The cost of a healthy diet is defined by FAO as the minimum cost of food, using the least expensive available items in each country, that people have to pay to secure a healthy diet. The latter is based on the cost of the average quantities of each food group recommended in national Food Based Dietary Guidelines (FBDG). On average, the healthy diet meets 95% of nutrient needs.

To determine affordability of a healthy diet, the cost of a healthy diet is compared with country-specific income distributions available in the "Poverty and Inequality Platform" (PIP) of the World Bank. From these data, the percentage and number of people in each country / region who cannot afford a healthy diet are estimated. The threshold for an unaffordable healthy diet in 2020 was when its cost exceeded 63% of the income in a country. However, food prices have surged since 2020, so this threshold for unaffordability is likely to change. For more details on the derivation of CAHD and the estimates of the number of people in the world and by region and subregion, see Chapters 2 and 3 and ANNEX 3 in FAO, IFAD, UNICEF, WFP and WHO (2022).

2.2.1b Food security indicators at national level

FAO (2006)has emphasized that "Food Security" is

> "a state in which all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life"

	Cost of a healthy diet in 2020		People unable to afford a healthy diet in 2020		
	Cost (USD per person per day)	Change between 2019 and 2020 (percent)	Percent	Total number (millions)	Change between 2019 and 2020 (percent)
WORLD	**3.54**	**3.3**	**42.0**	**3074.2**	**3.8**
AFRICA	**3.46**	**2.5**	**79.9**	**1031.0**	**2.5**
Northern Africa	3.57	-0.7	57.2	136.7	-0.8
Sub-Saharan Africa	3.44	2.9	85.0	894.3	3.1
Eastern Africa	3.37	3.4	87.4	360.8	3.0
Middle Africa	3.34	2.2	85.4	152.2	3.0
Southern Africa	3.84	3.3	65.5	44.2	1.8
Western Africa	3.45	2.7	85.7	337.1	3.3
ASIA	**3.72**	**4.0**	**43.5**	**1891.4**	**4.3**
Central Asia	3.11	4.0	21.5	7.5	6.9
Eastern Asia	4.72	6.0	11.0	174.4	18.7
South-eastern Asia	4.02	4.2	53.9	347.2	4.7
Southern Asia	3.81	4.0	70.0	1331.5	2.7
Western Asia	3.22	2.9	17.8	30.9	-1.4
LATIN AMERICA AND THE CARIBBEAN	**3.89**	**3.4**	**22.5**	**131.3**	**6.5**
Caribbean	4.23	4.1	52.0	13.9	3.5
Latin America	3.56	2.5	21.0	117.3	6.9
Central America	3.47	2.1	27.8	43.1	9.8
South America	3.61	2.7	18.4	74.2	5.3
Low-income countries	3.20	2.7	88.3	454.2	3.0
Lower-middle-income countries	3.70	2.9	69.4	2230.7	2.9
Upper-middle-income countries	3.76	2.9	15.2	374.0	10.9
High-income countries	3.35	4.0	1.4	15.3	3.3
OCEANIA	**3.07**	**3.6**	**2.7**	**0.7**	**1.0**
NORTHERN AMERICA AND EUROPE	**3.19**	**3.2**	**1.9**	**19.8**	**5.4**

Table 2.3. The cost of a healthy diet is the benchmark 2017 USD cost per person per day, projected forward to 2019 and 2020 using FAOSTAT data for each country's Consumer Price Index (CPI) for food, and World Development Indicator data for purchasing power parity (PPP) exchange rate. The people unable to afford a healthy diet is expressed as the weighted percentage (%) and the total number (millions) of the population in each region and country income group who could not afford the diet in 2020. For country income groups, the 2021 World Bank income classification is used for the years 2019 and 2020. From FAO, IFAD, UNICEF, WFP and WHO (2022).

Five indicators present in the FAO "Suite of indicators of food security" are derived from 3y moving averages of FBS data and are shown in Box 2.3. Note that none of these indicators provide information on the affordability, accessibility, or consumption by different population groups including nutritionally vulnerable groups. For more details, see: FAO (2022). The average dietary energy supply (DES) adequacy is defined as the dietary energy supply as a percentage of the average dietary energy requirement. The latter is defined as the amount of dietary energy measured in kilocalories that is required by an individual to maintain body functions, health, and normal activity. The estimates of DES adequacy consider the age structure, sex, and height in each country.

The dietary energy available from cereals, roots, and tubers (kcal/capita/day) (percent, 3y average) is an important nutritional

indicator because items from these food groups are generally the least expensive. As a result, these staple foods are often a large part of the diet in low-income countries. Consequently, the nutrient density, dietary diversity, and micronutrient supply of these diets are low as described in detail in Chapter 8c.

As noted earlier, average protein supply (g/capita/day, 3-year average) provides insight into the nutritional quality of the food supply at the global, regional, or national level.

The average supply of protein of animal origin (g/capita/day, 3-year average) is an additional important indicator of diet quality as animal-based proteins contains all the essential amino acids. The supply of protein from animal-based sources includes meat commodities, eggs, and dairy products, and fish and seafood, all measured in g/capita/day. Figure 2.4 presents the daily protein supply from both animal and plant-based foods in 2020. The data from 1961 to 2019 show that there has been an increase in the share of total protein from animal-based sources in most countries since 1961 (Roser et al., 2023).

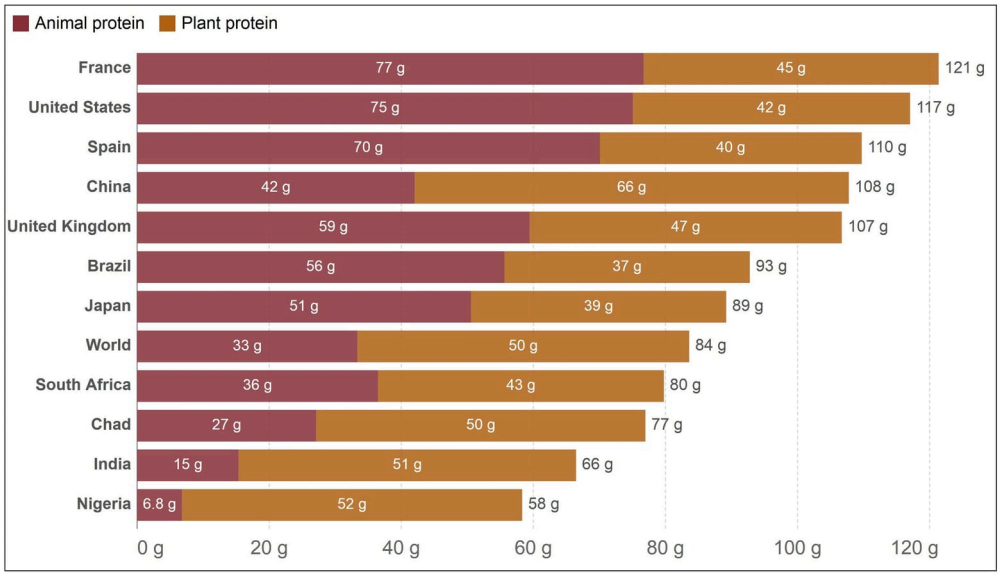


Figure 2.4. Daily protein supply from animal and plant-based foods (2020). Protein of animal origin includes protein from all meat commodities, eggs, and dairy products, and fish and seafood. Data (g/capita/d) from FAO. Map from ourworldindata.org / diet-composition. Additional information is available.

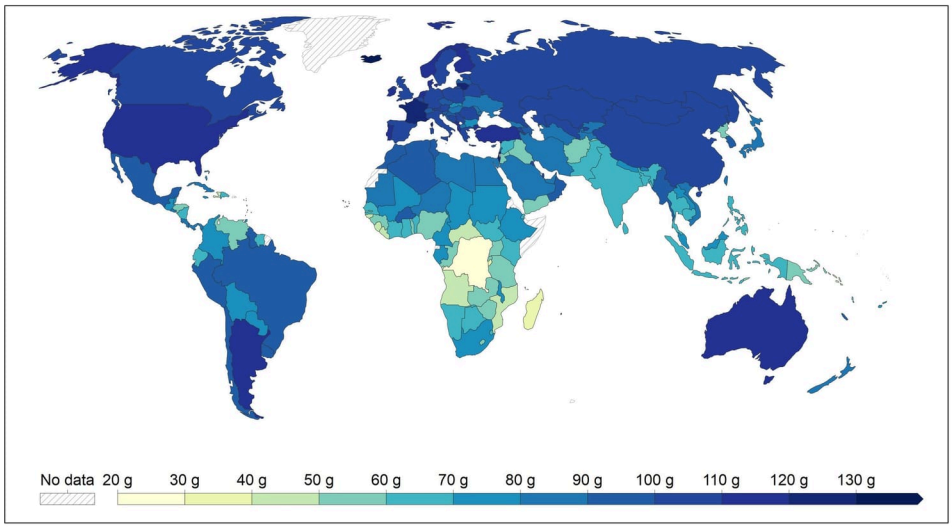
Additional indicators of the quality of the food supply based on FBS data have been developed by INDDEX in their Data4Diets platform. They include meat consumption (expressed as total kilograms per capita at the national level), national energy available from non-staples (i.e., all food items excluding grains and tubers), and national fruit and vegetable availability in the food supply. For more details, see INDDEX Project (2018).

2.2.1c Macronutrient availability at the national level

Macronutrients include carbohydrate, protein, and fats, and are the major source of energy in the diet. Macronutrient supply per capita can be estimated by combining supply data from FBSs expressed in g/capita/day with a food composition database for all primary and processed food products, as noted earlier. A common global food composition database is available from FAO and is used to estimate the nutrient supply data. The source of the nutrient composition data in this global database is not cited and any variation in the nutrient composition of food items across countries is not considered.

As shown in Table 2.1, the supply per capita for energy is expressed as kcal/capita/day, whereas the macronutrient supply per capita for both protein and fat are expressed as g/capita/day). Data are freely available by year and country from FAO (2021c).

Figure 2.5 presents the global per capita supply of protein and shows that in lower-income countries the per capita protein supply in 2020 is about 60-80g/capita/day, whereas in most countries across Europe, North America, and Oceania, the per capita protein supply is greater than 100g/capita/day.



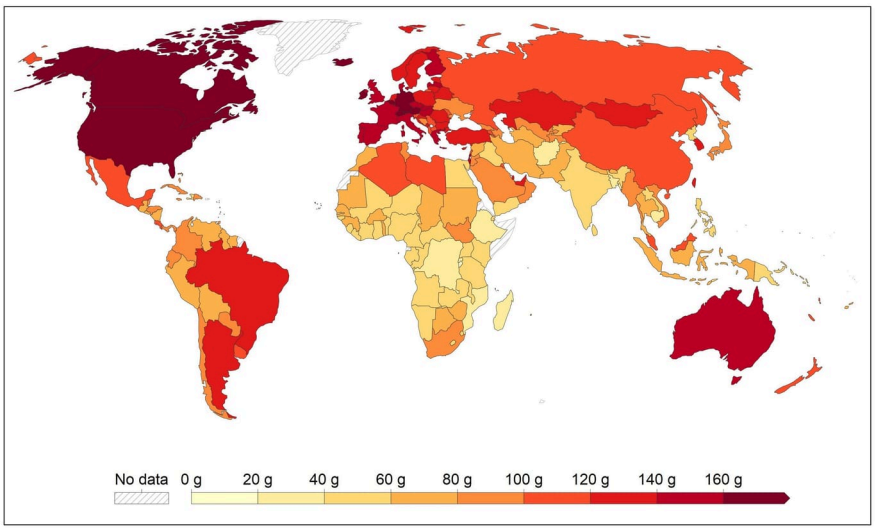
Note: Data measures the availability delivered to households but does not necessarily indicate the quantity of protein actually consumed (food may be wasted at the consumer level).

Figure 2.5. Global per capita protein supply (2020). Data from FAO. Map from ourworldindata.org/food-supply. Additional information is

available

The quality of the diet improves with the increased consumption of protein-rich foods. At the national level, the average protein supply provides an estimate of the protein supply per capita and thus insight into the nutritional quality of the national food supply. However, no information on the availability of protein supply in nutritionally vulnerable groups is provide by the FBS data.

In contrast to both energy as kcal/capita/day and protein g/capita/day, the per capita fat supply g/capita/day has increased across all regions during the period 1961 to 2014, although over the last decade, the increase has slowed in North America, Europe, and Oceania. Regional differences in supply for fat are larger than that for either energy or protein, with the average capita fat supply in North America being almost three times as large as in Africa in 2020 (Figure 2.6, Roser et al., 2023).



Note: Data measures the fat availability delivered to households but does not necessarily indicate the quantity of fat actually consumed (food may be wasted at the consumer level).

Figure 2.6. Global per capita fat supply (2020). Data from FAO. Map from ourworldindata.org/food-supply. Additional information is available

Macronutrients are sometimes expressed as percentage of total energy (e.g. percentage of kilocalories from protein or fat) due to uncertainties in the data and process used to develop FBS. Values expressed in this way (i.e., as nutrient density) are said to be much less influenced by age and sex in a population. Moreover, assessment of protein supply as protein density allows the quality of the diets to be compared across countries which have differing levels of accuracy of the food supply data, and thus differing levels of under / over estimation of total nutrient per capita supply. Nevertheless, even comparison of the values for percentage of energy as well as percentage of energy from protein and fat (and any other nutrient) derived from each food group across

countries can be jeopardized by a differing level of under / over estimation between food groups. In some cases, values for the ratios of polyunsaturated fatty acids to saturated fatty acids (P:S) and unsaturated fatty acids to saturated fatty acids (U:S) have also been used (Sasaki & Kesteloot, 1992) rather than per capita nutrient supply.

2.2.1d Micronutrient availability at national level

Efforts have been made by several investigators to use FBS for the assessment of the adequacy of micronutrient supply at the global level, in different regions of the world, and in individual countries. To accomplish this aim, FBS have been matched not only to energy, protein, and fat, but also to a full range of micronutrients. However, because the FAO FBS do not include data on the micronutrient supply per capita at the present time, investigators have compiled their own data on the micronutrient composition of the food commodities listed in the FAO FBS using data from food composition tables.

As noted earlier, FAO FBS provide only national-level estimates per capita. This precludes the assessment of per capita micronutrient supply that takes into account age and sex and the differing micronutrient needs of the national population. However, by applying data from estimates on the age and sex distributions of the national population in each country, the theoretical population mean requirement for each micronutrient at the national level can be estimated. From this, the percentage of the population at risk of inadequacy has been calculated, defined as the proportion of the population with micronutrient intakes below the theoretical average micronutrient requirement of the population. The population distribution of intakes for each micronutrient is inferred based on the assumed mean per capita micronutrient intake and coefficients of variation (CV) for each micronutrient (taken from the literature). The latter takes into account inter-individual variation in micronutrient intakes arising from the inequities which affect access and availability of certain foods. This approach of assessing risk of inadequacy is related to the Estimated Average Requirement (EAR) cut-point method and described in Chapter 8b for a population.

Several investigators have employed this approach to estimate risk of micronutrient inadequacy based on FBS data. Wuehler et al. (2005) were the first investigators to compare the adequacy of the food supply to meet the theoretical population mean requirement for zinc, based on FBS for 176 countries.

In this analysis, estimates of the absorbable zinc content of the national food supplies were derived from a global food composition database compiled by the investigators, with zinc absorption predicted using a model developed by the International Zinc Nutrition Consultative Group (IZiNCG). In this study, the investigators compared the absorbable zinc content of the food supply with the population theoretical mean *physiological*requirement to estimate the percentage of the population at risk of inadequate zinc intake. This earlier analysis was revised by Wessells & Brown in 2012. Figure 2.7 shows the estimated country-specific prevalence of inadequate zinc intake generated from this revised analysis. A companion article highlights the major sources of uncertainty in the analysis (Wessells et al., 2012). They also compared their estimated prevalence of dietary zinc inadequacy with the national prevalence of stunting in children less than 5y in 138 low- and middle-income countries (Figure 2.8).

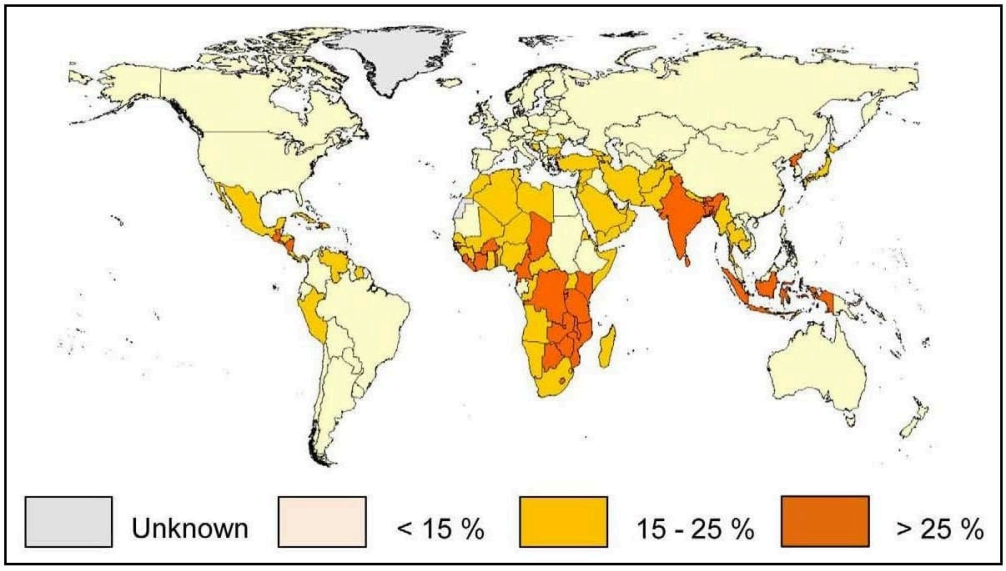


Figure 2.7. Estimated country-specific prevalence of inadequate zinc intake. Data are based on the composite nutrient composition database, IZINCG physiological requirements, the Miller Equation to estimate zinc absorption and an assumed 25% inter-individual variation in zinc intake. Data are for the 2005 time frame (2003-2007). Redrawn from Wessells & Brown (2012).

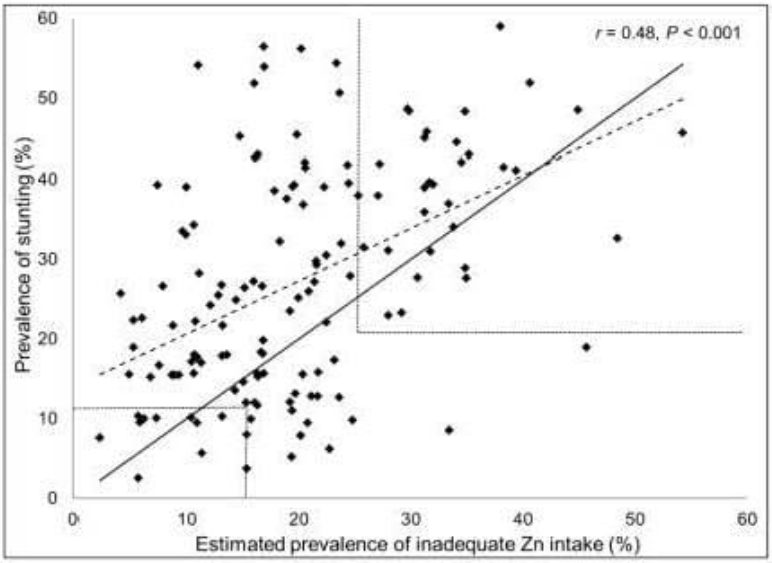


Figure 2.8 Relationship between the estimated prevalence of inadequate zinc intake and the prevalence of childhood stunting. Stunting data (low height-for-age) are for children less than 5y in138 low- and middle-income countries. The solid line represents the line of identity (intercept=0, slope=1). The dashed line represents the best-fit regression line. Dotted lines demarcate countries with a high risk of inadequate zinc intake and where the prevalence of stunting is > 20%. From Wessells & Brown (2012).

In later publications, this same approach has been applied to assess the per capita availability of several other micronutrients in regions (e.g., Africa) (Joy et al., 2014)and South Asia (Mark et al., 2016), as well as in specific countries (Broadly et al., 2012; Arsenault et al., 2015; Watts et al., 2015). For example, in a study by Arsenault and co-workers FBS data from three countries (Bangladesh, Senegal, and Cameroon) were used to assess the per capita availability for eight micronutrients (vitamin A, vitamin C, vitamin B_6, riboflavin, niacin, folate, calcium and zinc). Although a common food composition database was compiled and used for all three countries, additional country-specific adjustments were made in this study to take into consideration the in-country processing of staple foods, which lead to differences in the nutrient composition of ingested foods (e.g. percentage of refined wheat at country level). Two indicators of adequacy were assessed in this study:

- Per capita nutrient supply expressed as a percentage of the theoretical mean population-adjusted nutrient requirements for each country;

- Percentage of the population that had an adequate usual intake, based on the assumed inter-person variability in micronutrient intakes obtained from the literature.

Each nutrient gap in the food supply was also calculated as the difference between the current amount in the food supply minus the amount of the nutrient needed to achieve 80% prevalence of adequate intakes. In this study, linear optimization modeling was also used to determine the optimal mix of crops to fill the micronutrient gaps identified in each county and thus increase their micronutrient adequacy. Table 2.4 shows the optimal set of crops required to address the micronutrient gaps identified in Senegal while minimizing land requirements.

Table 2.4 Optimal set of crops to address micronutrient gaps in Senegal while minimizing land requirements[a]

a The amounts are additional amounts needed to achieve target levels of all nutrients so that the food supply provides sufficient amounts of nutrients (except calcium) for at least 80% prevalence of adequate intakes in the population

b Serving sizes are in edible form, i.e., all vegetables are cooked, grains are dry but nutrients adjusted for cooking losses. Data from Arsenault et al. (2015).

Crops	Servings[b]	Grams	Vit. A (µg RAE)	Vit. C (mg)	Ribofl. (mg)	Folate (µg)	Calcium (mg)	Zinc (mg)	% crop land
Barley	0.50	23	0	0	0.07	4	8	0.6	1.8
Broccoli	0.50	39	30	25	0.05	42	16	0.2	1.0
Cabbage	0.50	38	2	14	0.01	11	18	0.1	0.4
Carrots	0.40	18	157	1	0.01	3	6	0.1	0.3
Groundnuts	0.50	14	0	0	0.02	27	10	0.5	3.8
Okra	0.50	40	6	7	0.02	18	31	0.2	0.8
Pumpkin	0.25	31	90	1	0.02	3	5	0.1	0.4
Spinach	0.25	23	118	2	0.05	33	31	0.2	0.6
Total	3.40	226	403	50	0.25	141	123	1.8	9.0
%adequate intakes	—	—	81	%95	%85	%84	<1%	%80	—

In an effort to overcome some of the limitations of FBS data, notably their lack of information on both individual foods that are important in diets and actual intakes by age or sex, Smith et al. (2016) have constructed a new model: the Global Expanded Nutrient Supply (GENuS) model to estimate per capita supply for 23 individual nutrients across 225 food groups for 34 age-sex groups in 152 countries. They also provide trends in national level data over50 years (1961-2011).

To achieve these objectives, the GENuS model combines FBS data with some ancillary data from individual and household surveys at country level. This approach yields food supply estimates which consider both the edible weight of each food commodity and level of food fortification in 225 food groups. In each country an ad-hoc food composition database was compiled based on six national and regional food composition tables in which nutrient losses during the processing of cereals to flour were considered. Age and sex-specific supplies of foods and nutrients across 26 age groups were calculated for each country. The distribution of nutrient per capita supply by age and sex was derived through a probabilistic method (Monte Carlo simulations). Finally, to assess validity, a comparison of the GENuS estimates of nutrient supplies against USDA estimates was performed, which revealed very good agreement for 21 of 23 nutrients. For a discussion of the limitations of the data used in the GENuS model, see Smith et al. (2016).

This page titled 2.2: Food Balance Sheets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Household Surveys

Household surveys assess food available for consumption at the household level. This is defined as the total amount of food and beverages available for consumption in the household plus any foods prepared and consumed outside the household. Household surveys are of two types:

- Household Food Consumption Surveys (HFCS). These were developed in the 1980s as an alternative to surveys at the individual level, and provide measurements of food consumption in the household.
- Household Consumption and Expenditure Surveys (HCES). In these surveys, food available for consumption is measured. These household surveys are widely used to inform on global food security and poverty, particularly in low- and middle-income economies. Such surveys can also be conducted in institutions such as those for institutionalized elderly.

In the past, some surveys similar to HFCS or HCES were conducted in high-income countries with the objective of assessing food consumption at the household level. They were based on a record of food quantities entering the household, either purchased, received as gifts, or produced for household use, over a reference period, often seven days. They used the “food account method” Burk & Pao, 1976; or the “food inventory method&rdquo. (Turrini et al., 2001)

Food consumption by individual members of the household was often recorded as an additional component of the survey. This approach led to an increased burden on the sampled households, resulting in some cases in a low response rate. Such surveys are now rare and have been replaced by food consumption surveys at the individual level.

During HFCS and HCES, data on acquisition and/or expenditure and/or consumption of commodities including foods and beverages are collected. HFCS and HCES are less expensive than individual food consumption surveys and require a lower burden for participants. However, many of the technical problems and limitations of household surveys are the same as those of dietary surveys conducted at the individual level, which are discussed in detail in Chapters 4 and 5. For example, diet might be altered by the design of the survey or the recording process.

Typically, HFCS and HCES are conducted on a large nationally representative sample of households (Fiedler et al., 2012), and include the collection of information on demographic and socioeconomic characteristics of the household, thereby enabling data to be presented in terms of income level, family size, region of the country, etc. Attention must be paid to the sampling design of these surveys to ensure that a representative national sample is obtained; see details on sampling methods in Chapter 1, section 1.4.2. The sampling design should account for the influence of season, holidays, weekends, socioeconomic status, and region on food consumption patterns.

2.3.1. Household Food Consumption Surveys (HFCS)

Household Food Consumption Surveys measure all food and beverages consumed within a household during a specified period. They require careful supervision by the interviewer and good cooperation of the respondents. In general, these surveys are more complicated and costly to undertake than HCES and hence were often conducted infrequently. Such surveys are rarely performed today. Instead, they have been replaced by surveys of food consumption at the individual level which are more informative, even if more costly (see Chapter 3). Two methods — weighed food records and household 24hr recalls — were used in the past to measure the household food consumption.

Weighed household food records

For this method, the food eaten by the household is recorded using weighed food records (also called food diaries). These are usually completed over at least a 1wk period, by either the householder or a fieldworker. During the survey period, the weight or volume of each food consumed at each meal is recorded, before subdivision into individual helpings USDA. Detailed descriptions of all foods, including brand names, and their method of preparation are recorded. For composite dishes, the amount of each raw ingredient used in the recipe and the final weight of the prepared composite dish is also recorded. In some surveys the plate waste from each meal is collected and separated so that waste for individual food items can be weighed and recorded. Generally however, kitchen and plate waste, and food fed to pets, is not accounted for in this method, and instead an arbitrary wastage factor is applied.

Household 24-h recalls

In this method, the household member responsible for the food preparation is interviewed to obtain information on both household composition and household food consumption over the previous 24h period. In the first stage of the interview, information is collected on the dishes and ingredients consumed, followed by details on the quantity, focusing particularly on those foods that are important sources of energy. A technical guide for measuring household food consumption using a 24h recall was developed by Swindale & Ohri-Vachaspati (1999). This guide provides detailed instructions and sample questionnaires that can be used to collect the data, quantify the portion sizes of food consumed, and analyze the results.

2.3.2 Household Consumption and Expenditure Surveys

Household Consumption and Expenditure Surveys (HCES) are particularly important and typically performed within Household Expenditure Surveys (HES), i.e. economic surveys designed to inform national economic policy and usually implemented by national statistical agencies.

HCES are less costly than HFCS and are widely used in low-income settings (FAO, 2023). They are usually conducted every 3-5y.

HCES refer to a heterogeneous group of complex surveys which include Household Income Expenditure Surveys (HIES), National Household Budget Surveys (NHBS), and Living Standards Measurement Surveys (LSMS). The latter (i.e, LSMS) are multi-topic surveys for which technical assistance is usually provided by the World Bank's Living Standard Measurement Study (LSMS) group.

HCES were designed primarily to measure household food available for consumption in a reference period through the expenditure approach (i.e., the monetary value of the food). Over time, quantities of foods and beverages have also been collected in an effort to repurpose the surveys so they can serve other needs such as monitoring household-level food security. This has led to improvements in survey design to capture food data of a higher quality by including a more detailed food component module.

For some HCES based on the concept of acquisition, data are collected on food entering the households acquired through purchases, own-production, and in-kind, on the assumption that there were no major changes in household stocks during the survey period. An important limitation of this simplified methodology is related to the fact that some foods (e.g. grains) are not perishable and can be stored. Consequently, if the survey was conducted in a period of drawing-down stocks to meet current consumption, household consumption would be underestimated. On the contrary, if the survey was conducted in a period of accumulating stocks for later consumption, household consumption would be overestimated (Smith et al., 2014). As a result, use of these simplified HCES is no longer recommended. Alternatively, a combination of acquisition and consumption data can be collected in the HCES so that both food acquired through purchases and consumed from own-production and transfer are reported.

Food consumption estimates generated from acquisition data or a combination of both acquisition and consumption data are usually referred to as “apparent consumption” or “food available for consumption” to distinguish it from actual consumption (Fiedler & Mwangi, 2016). Data collected in HCES on food available for consumption are quite often used as a proxy for the quantities eaten.

HCES vary in their complexity and respondent burden. Data may be collected through a food diary, recall, or a list, with a reference period usually ranging from seven days to one month. When a list is used, the food and food groups specified varies, and when the list is short, the aggregated food items often lack the details needed to match the food items correctly with a food composition database. Food and beverages wasted, spoiled, or fed to pets or livestock are frequently not taken into account, although in some cases, waste may be weighed or a wastage factor applied.

In general, in the past the capture of food and beverages prepared and consumed outside the home has also been poor in most HCES, although there is now a trend to improve the recording of the consumption of these items. For example, in some recent HCES a special module has been added to collect personal expenditure on snacks, meals, sweets, and drinks consumed outside the home. Adjustments are also made in some HCES for the presence of non-household members during the survey period. Ideally HCES should report all sources of food consumption, either purchased, received as gifts, or produced for household use.

Variations in the survey designs of HCES across countries and over time within countries has created many challenges. To understand the implications of these variations on the accuracy and consistency of consumption measurements, a field study was conducted in Tanzania in 2007 and 2008 in which consumption estimates in 4,000 households was measured via eight different HCES survey designs, selected to capture the most common designs used in practice (Beegle et al., 2010). The effect of variations according to survey length, use of a household diary versus recall format, and level of detail of food items listed in the HCES on over- and under-reporting was investigated, using frequently supervised personal consumption diaries as the “gold standard”.

The characteristics of the study population, notably whether they were illiterate or largely urban, was found to affect the magnitude of the underestimate of consumption using the HCES diary method. For the HCES recall method, the investigators recommended the use of a full long list of food commodities to recall within a reference period of 1 or 2 weeks rather than using a shorter recall list with few aggregated food categories. In summary, over- and under-reporting appeared to depend on the specific setting and varied with traits such as the literacy of the household, the share of consumption from home production, dietary diversity, and fraction of meals eaten in the household as opposed to consumption by individuals at restaurants and catering establishments (Beegle et al., 2010).

The findings of Beegle et al. (2010) highlighted the need to improve the quality and utility of the food data components of the HCES surveys so they were more relevant to nutritionists and food security analysts, as well as the importance of harmonizing the HCES survey methodology. Consequently, a desk review of survey questionnaires and methods used in 100 available HCES surveys from low- and middle-income countries was conducted to identify key areas for improvement (Smith et al., 2014).

Box 2.4 Areas identified in the food data component of HCES surveys warranting further investigation

- Choice of a diary or recall survey, and the appropriate reference period
- Food consumed away from home and cooked/packaged meals
- Measuring individual versus household consumption
- Measuring food acquisition versus food consumption
- Length and specificity of survey food lists

From Zezza et al. (2017).

Five areas in the food data component of HCES surveys were selected for investigation; these are summarized in Box 2.4. For a detailed discussion of the measurement error patterns that were identified within each of these five areas, see Zezza et al. (2017). As an example, a major challenge has been the rise in the consumption of food away from home, a trend that has accompanied the nutrition transition over the last decades in both developing and developed countries. Such data may include food that is purchased (e.g., from a street stall or restaurant), and/or received in kind (e.g., provided via food assistance or a gift etc). Snacks have also become an increasingly important part of the diet and may make up a large proportion of food consumed away from home.

Food away from home (FAFH) can refer to food produced outside the home such as takeout meals, irrespective of whether the food is consumed outside or inside the home. Alternatively, food away from home can refer to food consumed outside regardless of the origin of the food. Examples of this scenario includes homemade meals consumed at work or school. The FAO and the World Bank 2018 have outlined a method for defining food away from home, as shown in Figure 2.9.

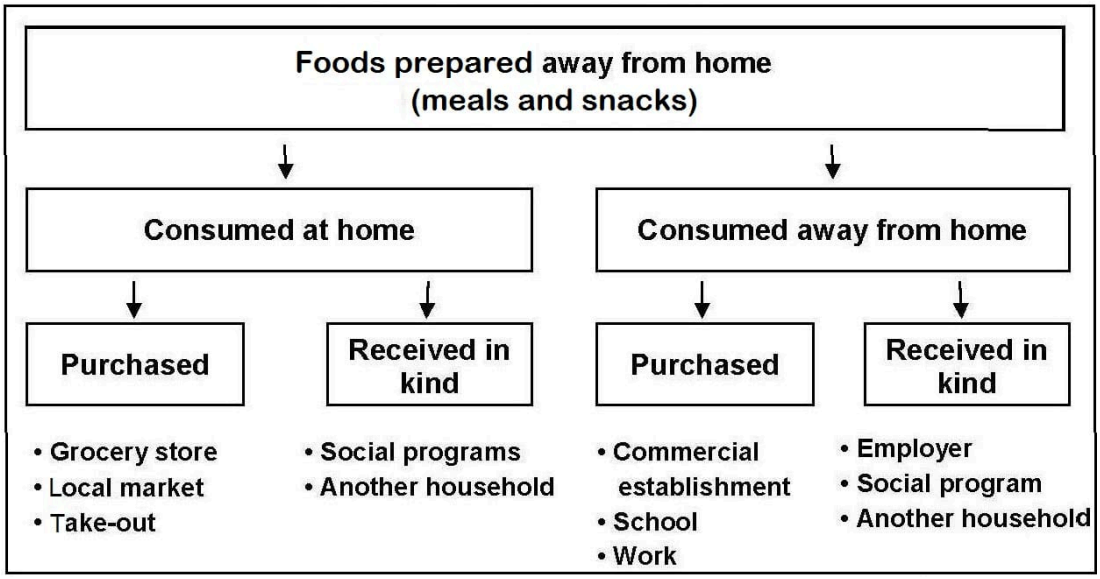


Figure 2.9 The definition of food away from home (FAFH). Redrawn from FAO and The World Bank (2018).

For nutrition and food security analysis, an additional challenge, not captured in Figure 2.9, is acquiring information on what was eaten. Such details must often be secured from other sources. Several low- and middle-income countries have developed innovative approaches to secure these details; see Annex 2 in FAO/World Bank (2018) for more details.

Table 2.5 presents data on food consumed away from home based on data collected from the 100 HCES surveys examined by Smith et al. (2014). Such data may include food that is purchased (e.g., from a street stall or restaurant) and/or received in kind (e.g., provided via food assistance or a gift etc). Snacks have also become an increasingly important part of the diet and can make up a large proportion of food consumed away from home. As shown in Table 2.5, although most of the surveys collected data on food consumed away from home, the detail with which the data were collected was very poor.

Table 2.5 Food away from home data collection. [a] N=100 surveys. [b] Calculations are only for surveys for which any data were collected on food consumed away from home (N=90). Data from Smith et al. (2014).

	Interview surveys (%)	Diary surveys (%)	All (%)
Whether any data collected on food consumed away from home[a]	83.3	100.0	90.0
Detail of data collection[b]			
Only one line item (e.g., “Restaurant food”)	36.0	7.9	23.9
Data collected for multiple places of consumption	14.0	35.0	23.3
Data collected on food received in-kind	46.0	65.0	54.4
Data collected on specific food items	28.0	40.0	32.9
Snacks explicitly referred to	26.0	35.1	29.9

	Interview surveys (%)	Diary surveys (%)	All (%)
Alcoholic beverages explicitly referred to	36.0	32.4	34.5
Data collected at the individual level	12.0	23.7	17.0
Recall period [b]			
Less than one week	60	100.0	47.8
One week	48.0	0.0	26.7
Two weeks	12.0	0.0	6.7
One month	14.0	0.0	7.8
Greater than one month	20.0	0.0	11.1

In view of the many challenges experienced in collecting the food data component of HCES surveys, and the recommendations proposed by several investigators (Carletto, Zezza & Banerjee, 2013; Smith et al., 2014; Zezza et al., 2017), guidelines to improve the quality of data collected have been developed by FAO and the World Bank (2018) and are itemized in Box 2.5.

Box 2.5 Domains and their accompanying recommendations

Recall versus diary and length of reference period

- Low-income countries are advised to adopt recall interviews and a seven-day period

Seasonality, number of visits

- Conduct one visit per household, spreading the sample over 12 months of fieldwork or
- Conduct two visits per household, when the timing of the visits is scheduled to capture seasonal variation

Acquisition versus consumption

- All surveys should collect data on all main modes of food acquisition, namely: purchases; household's own production; received in kind
- For consumption, it should be clear whether the questions concern food intended for consumption (including food waste) or food actually consumed (net of food waste).

Meal participation

- Preferably include an individual household member-based meal module.

Food away from home

- Design a separate module, based on a clear definition of food away from home, with data collection around meal events including snacks and drinks

List of food items

- Preferably adopt a standard food classification system for survey harmonisation e.g., food groups for COICOP classification list

Non-standard units of measurement

- Report on standard units and non-standard units with related conversion factors, where feasible for households

COICOP, Classification of Individual Consumption According to Purpose
From FAO and WORLD BANK (2018).

These guidelines take into consideration the best balance between accuracy and cost effectiveness. For more details on the recommendations specified for each domain and their justification, see FAO and World Bank (2018). Low- and middle-income countries are recommended to adopt these guidelines to improve the quality of the food data component collected in HCES and standardize the methodology across countries.

2.3.3 Uses of household surveys

Increasingly, HCES are being repurposed so that the data generated are not only used by economists for monitoring poverty, calculating national accounts, and as an input for consumer price indices, but also by food security analysts and nutritionists for nutrition-related analyses, based on the food consumption module of the HCES surveys (Smith et al., 2014; FAO and the World Bank, 2018). The World Bank has developed the most comprehensive repository of publicly available worldwide HCES data. An additional resource on HCES surveys is also available at the International Household Survey Network (HSN).

Several resources are available to derive food security and nutrition indicators as well as nutrient-based indicators based on the food module of HCES surveys. Of these, two resources itemized below provide details of their methods of construction, uses, strengths and weaknesses.

- Data4Diet platform compiled by the INDDEX project (2018) is freely available at: Data4Diet.
- ADePT-FSM Food Security Module, a stand-alone application, based on the research of Moltedo et al. (2018) and developed by the World Bank and FAO.

Readers are advised to consult these two resources. A selection of indicators designed to assess food security and nutrition are summarized in Box 2.6, whereas details of indicators designed to characterize dietary micronutrient supply are presented in Box 2.7. When compiling these indicators, the term “apparent consumption of foods” and “nutrient supply” at the household level and “apparent nutrient intake” at the individual level rather than “intake” should be used. None of these indicators permit assessment at the individual level.

Box 2.6 Selected Indicators of Food Security and Nutrition

- **Prevalence of Undernourishment** (PoU percent) is used by FAO to monitor the state of food insecurity in the world and is published annually in the State of Food Insecurity Report (SOFI), by FAO, IFAD, UNICEF, WFP and WHO (2022). Updated instructions related to the estimation of PoU using HCES data are available: FAO e-learning course
- **Depth of Food Deficit** (kcal/capita) represents the average per capita amount of additional energy (kcals) needed for undernourished individuals to meet the Average Dietary Energy Requirement (ADER). The Depth of Food Deficit is used to understand the degree of food insecurity in a country and compare the magnitude across countries. It does not measure dietary quality. For this estimate, the FAO probabilistc cut-point method is applied; see Moltedo et al. (2022) for more details. Cafiero et al. (2014) discusses the validity and reliability of this and other food security measurements.
- **Balanced Diet Indicator** is a measure of dietary quality and defined as the proportion of population with access to a balanced diet. For this, households are classified according to whether their dietary energy consumption from sources of protein, fat, and carbohydrate are below, within, or above recommended thresholds (i.e., protein, 10-15%; fat, 15-30%; total carbohydrate, 55-75%). When significant proportions of the population fall outside these thresholds, then there is a possibility of adverse consequences associated with chronic diseases within a country.
- **Household Adequacy of Fruit and Vegetable Consumption** expressed as g per capita per day, provides a measure of diet quality and an understanding of dietary patterns. It is used to identify inadequacy of fruit and vegetable consumption across countries and for subgroups based on household income, sex of head of household, geographical area etc. For this indicator, the total weight in grams of fruit and vegetables consumed by the household is divided first by the number of household members, followed by the days of survey reference period.

$$\begin{aligned} \geq 400 \text{ g/ capita/day (i.e., 5 servings)} &= \text{adequate} \\ < 400 \text{ g/ capita/day} &= \text{inadequate} \end{aligned}$$

Table 2.6 presents a comparison on apparent fruit and vegetable consumption (edible g/capita/day) from HCES data by income quintile in five countries and shows that apparent fruit and vegetable consumption only surpassed 400g/capita/day for income quintiles 3, 4, and 5 for Country 2.

- **Proportion of animal protein** (as percent of total protein). If details are available, this measure can be used as a proxy of dietary quality at the population level. Higher percentage of animal source protein likely to be associated with a higher absolute consumption of animal source foods that provide a variety of micronutrients. The latter are either less abundant or less bioavailable in plant-based food sources. Food commodities considered as animal source foods include meat, fish, eggs, milk, and cheese. Calculated as:

$$\%\ \text{animal protein} = \frac{\text{Total protein animal sources (g)} \times 100}{\text{Total protein (g) from all foods}}$$

Table 2.6 Fruit and vegetable apparent consumption by income quintiles in five countries. From ADePT-FSM software

	Fruit and vegetable consumption (Income Quintile — edible g/capita/day)				
	1 (Lowest)	2	3	4	5 (Highest)
Country 2	220	360	471	603	674
Country 3	123	148	163	194	258
Country 4	127	202	248	295	349
Country 5	116	173	173	205	261
Country 6	58	75	86	98	119

See ADePT-FSM SOFTWARE for more details.

Indicators designed to characterize household dietary nutrient supply Additional indicators designed to quantify, characterize, and evaluate the adequacy of household micronutrient supply across populations have also been compiled. These indicators are calculated by first multiplying the edible amounts of the food commodities listed in the food supply module of the HCES survey and said to be available for consumption (not the actual intake) for each household, with their corresponding nutrient content derived from food composition tables (FCTs) (mostly for the raw form of the food before preparation). The data generated are termed “micronutrient supply”. The nutrients included are the essential amino acids, vitamins B_1, B_2, B_6, B_{12}, and C, folate, total vitamin A (expressed as both Retinal Equivalents and Retinol Activity Equivalents), zinc, calcium, and total iron. In addition, heme iron, expressed as a percentage of total iron, can also be estimated.

Several investigators have compared the estimates of apparent nutrient intake derived from HCES data with those obtained with individual dietary assessment methods. In a systematic review of five studies, Tang et al. (2022) concluded that in general nutrient intakes were overestimated using HCES food consumption data. For some of the studies, discrepancies were large compared with the estimates from individual dietary assessment methods. For a discussion of the possible reasons for the poor agreement, see Tang et al. (2022).

Apparent micronutrient intake per capita per day and per adult-male equivalent per day as well as micronutrient density can be estimated at the national level, and disaggregated into categories when they can be used to monitor or identify target populations for policy makers. The categories that can be included depends on the household socio-economic characteristics collected in the HCES in the country and may include education, income, geography, and ruralness. Examples of these indicators disaggregated by region, urban-rural areas, and quintile of income are given in Box 2.7.

Box 2.7 Examples of indicators* based on average apparent micronutrient intake at the individual level

- **Apparent zinc intake** (mg/capita/day) at national, income quintile, and urban-rural levels. Note that the apparent zinc intake is not adjusted for bioavailability in view of the paucity of phytate data currently available in FCTs

- **Zinc density** (mg/1000kcal) at national, income quintile, and urban-rural levels. Density computed as mean of the nutrient density of the diet of the household
- **Apparent folate intake** (expressed as Dietary Folate Equivalents — µg/capita/day) at national, income quintile, and urban-rural levels
- **Dietary Folate Equivalent density** (ug/1000kcal) at national, income quintile, and urban-rural levels
- **Apparent vitamin A intake** expressed as Retinol Eqivalents (µg RE/capita/day) or Retinol Activity Equivalents (µg RAE/capita/day) at national, income quintile, and urban-rural levels
- **Retinol Activity Equivalent density** (µg RAE/1000kcal) at national, income quintile, and urban-rural levels
- **Apparent total iron intake** (mg/capita/day) at national, income quintile, and urban-rural levels
- **Apparent heme iron intake** (as percent of total iron) at national, income quintile, and urban-rural levels. Heme iron content estimated from raw and cooked meat from literature sources
- **Prevalence of Nutrient Inadequacy (PoNI)** for vitamins B_1,B_2,B_6, B_{12}, vitamin A, vitamin C, Ca, Zn. Estimates via probabilistic EAR cut-point method.

* Depending on the household characteristics collected in the HCES data, national and regional data can also be disaggregated by category of socio-economic group, education, household size, etc.

In the past, attempts to assess micronutrient adequacy from HCES survey data estimated ratios of apparent micronutrient intake per capita to the average weighted micronutrient requirement at the population level. These ratios provided a gross indication of issues with regard to meeting requirements of selected micronutrients at the population level, but provided no information on the prevalence of inadequate micronutrient intakes in the population (Moltedo et al., 2018).

Consequently, recently efforts have been made to estimate the prevalence of nutrient inadequacy (PoNI) in the population based on HCES survey data. The new method extends the FAO probabilistic cut-point method used earlier to estimate chronic dietary energy inadequacy to include micronutrients. To achieve this objective, the data generated on apparent micronutrient intake per day at the individual level must first be pre-adjusted to yield the distribution of apparent usual micronutrient intakes by removing the excess variation due to both within-person day-to-day variability and seasonal variability. After these preadjustments, the distribution of apparent usual intake levels is compared to a threshold to estimate PoNI. The threshold is based on the weighted average of the estimated average requirement (EAR) for the micronutrient under study for each sex-age group in the population.

The prevalence of inadequacy for eight micronutrients: vitamins A, B_1, B_2, B_6, B_{12}, and C, and calcium and zinc based on pre-treated household survey data versus individual level dietary data from Bangladesh has been compared by Moltedo et al. (2022). Their estimates of inadequacy from the household survey data compared with the individual-level data in which the conventional EAR cut-point method was applied, appeared promising. Nevertheless, more research is needed to guide the use of HCES to estimate apparent nutrient intake at the individual level and prevalence of inadequacy. Detailed instructions for estimating these indictors based on nutrient analysis from FCTs and the new FAO probabilistic cut-point method, respectively, are available at:

- Upgrading ADePT-FSM for analyzing food security statistics using household consumption and expenditure survey data
- Moltedo et al. (2022)

Special attention must also be given to the choice of the food composition tables used for the micronutrient analysis. For discussion of some of the limitations of the micronutrient values in food composition tables, consult Section 1.2 in the above ADePT-FSM Version 3 Software. Note that this version of the ADePT software also permits an assessment of the micronutrient content in foods consumed away from home.

This page titled 2.3: Household Surveys is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)

Both FBS and HCES are prone to measurement errors, as itemized in Table 2.7. For more details, see Thar et al. (2020) and FAO (2001) for FBS, and Fiedler et al. (2012) and Smith et al. (2014) for HCES.

Table 2.7. Comparison of Food Balance Sheets and Household Consumption and Expenditure Surveys. From FAO (2022a).

	Food Balance Sheets	Household Consumption and Expenditure Surveys
Objective	Assessment of food available for consumption at national level	Assessment of food available for consumption at the household level
Advantages	• Not expensive • Publicly available for over 245 countries & territories from 1961 • Snapshot of overall agri-food situation • Updated regularly with standardized methodology so compatible for comparison and trends of per capita food supplies in terms of quantity, caloric value, protein & fat content	• HCESs conducted regularly & publicly available • More detailed in sub-national, sex, economic strata breakdown • Fortfiable foods can be identified from percentage of HHs consuming & purchasing individual food • Report metrics that characterise food security, dietary quality, household nutrient supply, & apparent nutrient intake at individual level
Disadvantages	• Quality of FBSs limited by completeness and accuracy of in-country reports; data on country's stocks not always available • Food loss, food waste, homegrown food, & food fed to animals is under-recorded • Number of migrants & tourists not always subtracted from population estimates • Wrong matching of foods with FCTs as ambiguous details of foods listed • Nutrient supply data estimated from a common global FCT so differences in composition of foods across countries not considered • Distribution of food at sub-national level or population specific characteristics such as age, sex, SES not provided so high risk groups cannot be identified • No data on home-grown food or details of processed foods	• Relatively expensive & requires institutional capacity & trained personnel • Household food supply module not yet standardized across countries • Individual data on food purchases, food consumed from home production, food received in kind, and details of food consumed away from home not always collected • At-home food list varies (50-300 items) & not always specific leading to incorrect matching with FCTs • Seasonality of food consumption patterns not always considered • Food given to non-HH members not always reported • HCESs do not provide information on distribution of foods consumed among household members or by life-stage groups • Apparent male equivalent intake estimates limited by lack of data on physical activity, physiological state of household members

Grünberger (2014) compared per capita food supply based on country-specific FBS with that of 64 HCESs from 51 low- and middle-income countries, for 16 major food groups. Overall, the results suggested that the estimated differences in the average total food supply per capita are moderate. However, underestimates for the contribution of the average consumption of cereals, eggs, fish products, pulses and vegetables are likely, whereas overestimates for fruits, meat, milk, and sugar products are probable. These findings suggest that considerable uncertainty may exist in the estimates for the availability for consumption of single food groups at country and global level.

This page titled 2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

2.5: National and household surveys in the global assessment of food security and diet quality by age and sex

Where feasible, it is preferable to use quantitative food consumption data collected at the *individual level* to assess the distribution of usual intakes of both food and nutrients in different population groups within a country. In this way, the adequacy of the nutrient intakes at the population level can be evaluated using the methods described in Chapter 8b. Such quantitative data from food consumption surveys collected at the individual level can be accessed from the FAO/WHO GIFT platform.

However, currently, quantitative national food consumption data collected at the individual level are only available for a few low- and middle-income countries. Therefore, to fill this gap, alternative global databases based on information synthesized from FBS, HCES surveys, and individual-level surveys (where possible) have been compiled. Such databases are aimed at modeling worldwide individual intakes of foods and nutrients by age, sex and physiological status. Examples of these global databases include the Global Dietary Database (GDD) and a database developed by the Institute for Health Metrics and Evaluation (IHME) discussed briefly below. For more discussion of these global databases, see Chapter 3.

The GDD database provides modeled data in 188 countries for 55 dietary factors including 14 foods, 7 beverages, 15 macronutrients, 19 micronutrients, and two indices of carbohydrate quality. The FoodEx2 categorization system is used to standardize the description and classification of foods into food groups. Modeled data are provided by country, year of primary data collection, age across the life span, sex, education level, urban or rural residence, and pregnancy or nursing status. The GDD modeled estimates can also be used to compile global patterns of healthy and unhealthy diets together with indicators such as the Healthy Eating Index and the Mediterranean Diet Score, Minimum Dietary Diversity for Women (MDDW) and for Infant and Young Child Feeding (IYCF). See Chapter 8c for more details of these indicators.

The IHME initiative database aims to provide rigorous and comparable measurements of the world's most important health problems and evaluates strategies used to prevent them. IHME uses FAO Food Balance Sheet estimates, national product sales, household surveys, and data based on 24 hr recalls (considered the gold standard). Datasets created by IHME are stored in the IHME data catalogue known as the Global Health Data Exchange and can be freely downloaded from the IHME website. The IHME initiative provides modeled data by country, age, sex and year, based on primary data collected in 204 countries and 87 indicators, including 15 dietary indicators (9 foods and 6 nutrients). These dietary indicators are included in the Global Burden of Diseases (GBD), a worldwide observational epidemiological study that tracks the progress within and between countries of the changing health challenges (Lim SS et al. 2012; GBD 2017 Diet Collaborators). For example, Qiao et al. (2022) investigated the global burden of non-communicable diseases attributable to dietary risks from 1990-2019. They reported that a high intake of sodium and low intake of whole grains and fruits were leading dietary risks for deaths and disability-adjusted life-years (DALYs) worldwide, especially in developing countries and among males. Their findings highlight the need to raise public awareness of interventions and improve dietary practices aimed to reduce the disease burden caused by dietary risk factors. Data from the GBD study are also used to monitor progress in countries at the national and subnational level towards meeting the United Nations 33 health-related Sustainable Development Goals. Key research papers based on the global health estimates from the analysis of GBD data are published each year in a special issue of The Lancet.

This page titled 2.5: National and household surveys in the global assessment of food security and diet quality by age and sex is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

3: FOOD CONSUMPTION OF INDIVIDUALS

3.1: Methods for measuring food consumption of individuals
3.2: Technical improvements in food consumption measurements
3.3: Selecting an appropriate method

This page titled 3: Food consumption of individuals is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

3.1: Methods for measuring food consumption of individuals

This chapter describes methods commonly used for measuring the food consumption of individuals. Subsequent chapters discuss the factors associated with the reproducibility and validity of each of these methods (Chapters 5-7) and the calculation and, subsequently, evaluation of nutrient intakes (Chapters 4 and 8b).

Two groups of methods are used to measure the food consumption of individuals. The first group, known as quantitative daily consumption methods, consists of recalls or records designed to measure the quantity of the individual foods consumed over a one day period. By increasing the number of measurement days, quantitative estimates of the usual intakes of individuals can be obtained, using the same methods. The number, selection, and spacing of the days depend on the food intake, the nutrients of interest, the day-to-day variation in nutrient intake, and the level of precision required. Determination of the usual intake of individuals is particularly critical when relationships between diet and biological parameters or health or chronic disease are assessed. Estimates of usual intakes of individuals in a group or population are also needed to estimate the prevalence of inadequate intakes.

The second group of methods includes the dietary history and the food frequency questionnaire. Both obtain retrospective information on the patterns of food use during a longer, less precisely defined time period. Such methods can be used to assess the usual intake of foods or specific classes of foods. With modification , they can also provide data on usual nutrient intakes of individuals.

Quantitative daily consumption methods such as 24h recalls and records, when adjusted appropriately statistically to assess usual intakes, provide less-biased estimates of dietary intake than those such as food frequency questionnaires and dietary histories that are designed to generate data on usual intakes directly (National Cancer Institute, 2015).

The measurement of food consumption at the individual level is costly and time consuming. Hence, such studies should be planned with care (Thompson et al., 2015). Even with the advent of new technologies to measure dietary intakes (Section 3.2), challenges still remain. Consideration should be given to the cost-effective collection of additional data from the same individuals at the same time; such additional information may significantly enhance the interpretation of the dietary data. At a minimum, socioeconomic and health-related information, simple anthropometric measures, possibly a physical activity questionnaire, and biological samples for the determination of important biomarkers (Chapter 15) should be collected when time and resources permit (Buzzard and Sievert, 1994).

The accurate assessment of the food intake of infants is particularly difficult, especially when infants are receiving both breast milk and complementary foods (Piwoz et al., 1995). WHO (2023) has published guidelines that can be used to evaluate nutrient intakes of breastfed infants receiving complementary foods.

3.1.1 Twenty-four-hour recall method

In the 24h recall method, the respondent and, where necessary, their parents or caretakers are asked by the nutritionist, who has been trained in interviewing techniques, to recall the respondent's exact food intake during the previous 24h period or preceding day. Thus the method assesses the actual intake of individuals. However, a single 24h recall is not sufficient to describe an individual's *usual* intake of food and nutrients; multiple 24h recalls on the same individual over several days are required to achieve this objective (Section 3.1.2). Nevertheless, multiple single-day recalls on different individuals can give a valid measure of the usual intake of a group or population (Section 3.3.1).

A multiple-pass interviewing technique is recommended for the 24h recall methods. The USDA has developed an automated multiple-pass method (AMPM) consisting of five steps to collect dietary data using the 24h recall method; these five steps are shown in Table 3.1. The AMPM is used by the US in the National Health and Nutrition Examination Surveys (NHANES). Branded food products have now been added to the NHANES generic database to facilitate easier selection of foods consumed by the participants.

Table 3.1 USDA's Automated Multiple Pass Method. From (AMPM., 2013).

Design	AMPM Step	Memory Cues
Unstructured Respondents use their own strategy	**Quick List** ⇓ Collects recalled foods	Yesterday, midnight to midnight, day of week, activities, snacks and beverages, location
Structured Questions for specific types of foods	**Forgotten Foods**⇓ Probes for categories of forgotten foods	Seven question: Beverages, alcoholic beverages; sweets; savory snacks; fruits, vegetables or cheese; breads or rolls; and one for anything else.
Structured Times and names of eating occasions	**Time & Occasion**⇓ Collects eating times and meal names	Breakfast, lunch, dinner, snack
Structured Standardized questions to collect details of each food Review foods for each eating occasion & intervals between each occasion	**Detail Cycle**⇓ Collect food details and amounts.	For breakfast you had a bagel & coffee Anything else? Did you have anything between your 7am breakfast and 10am snack?
Unstructured Respondents use their own strategy	**Final Probe**⇓ Anything else recalled	Situations where foods may have been easily forgotten: in the car, at meetings, when shopping, cooking, or cleaning up?

The first step in the AMPM is the Quick List where respondents list all food consumed in the previous 24h period. The second step, called Forgotten Foods, includes a series of questions that probes for categories of foods that are commonly forgotten. The third step, Time & Occasion, collects the time each food was eaten and the name of the eating occasion. The fourth step is the Detail Cycle which elicits descriptions of foods and amounts eaten, aided by the interactive use of the USDA Food Model Booklet and measuring guide (USDA 2002+); see Chapter 5 for more details. The Detail Cycle also includes questions that review each eating occasion and each interval between eating occasions. The fifth step is a final review question, the Final Probe, which provides the respondent a final opportunity to recall any foods that had not been reported previously during the interview. The AMPM has been validated by comparing reported energy intake in young children with total energy expenditure assessed using the doubly labeled water method (Johnson et al., 1996) and in an observational study in which actual intakes of energy, protein, carbohydrate, and fat were compared with recalled intakes in men (Conway et al., 2004).

A modification of the multiple-pass 24h recall consisting of four passes — termed an interactive 24h recall — has been developed to collect information on rural populations in low- and middle-income countries; details are given in the technical monograph by Gibson and Ferguson (2008).

In the first pass a complete list of all the foods and beverages consumed during the preceding day is obtained.

In the second pass, a detailed description of each food and beverage consumed, including cooking methods and brand names (if possible) is collected. Standardized probe questions, specific for the population subgroup and setting, should be used to elicit specific details for each food item. For example, for milk products, probe questions should include the kind of dairy product, brand name (if appropriate), and percentage fat (as butterfat or milk fat). Further examples of probes that can be used to obtain detailed descriptions of specified foods are also provided in the monograph.

In the third pass, estimates of the amount of each food and beverage item consumed are obtained, generally in household measures, and entered either on the data sheet Appendix 3.1 or a computer-based data-entry form. Care must be taken to record whether any fortified foods or beverages have been consumed. Graduated photographs (Vossenaar et al., 2020), a set of measuring cups, spoons, and rulers, local household utensils (calibrated for use), play dough, or food models of various types (Chapter 5) can be used to assist the respondent in assessing portion sizes of food items consumed (Gibson and Ferguson 2008; Lazarte et al., 2012). See guidelines on portion-size estimation methods by Vossenaar et al. (2020). Vossenaar et al. (2022) also provide technical guidance on how to collect, compile, and use portion size estimation method conversion factors in a 24h recall. Information on the amount of any mixed dishes consumed by the respondents, the amount of each ingredient in the mixed dishes, and the total amount of each cooked mixed dish prepared must also be collected at this time. These details are usually recorded on a separate data sheet or computer-based data-entry “recipe form". Vossenaar et al. (2023) have published guidance on the use of standard and non-standard recipes in quantitative 24h dietary recall surveys.

In the fourth pass, the recall is reviewed to ensure that all items, including the use of vitamin and mineral supplements, have been recorded correctly. Methods for coding the completed 24h recalls and potential sources of coding errors are discussed in Chapter 5.

Additional modifications that can be used in the interactive 24h recall are listed in Table 3.2 and are discussed in more detail in the technical monograph (Gibson and Ferguson 2008). Whenever possible, recall interviews in rural settings in low and middle-income countries should be conducted in the respondent's home, because the familiar environment encourages participation, improves the recall of foods consumed, and facilitates calibration of local household utensils by the interviewer. For a review of potential measurement errors using self-reported 24h recalls in low-income countries and strategies for their prevention, see Gibson et al. (2017). However, major challenges arise when assessing food consumption at the individual level in countries or cultures where the majority of the food consumed is as shared plate eating. Strategies used to overcome some of these challenges are reviewed by (Burrows et al. 2019).

Table 3.2: Interactive 24-h recall modifications suggested for rural populations in developing countries to improve the recall of food items. From Gibson and Ferguson (2008).

Provide group training on portion size estimation before the actual recall.
Supply picture charts on the day before the recall for use as a checklist on the day the food is actually consumed, and for comparison with the recall to reduce memory lapses.
Provide bowls and plates for use on the recall days to help the respondents visualize the amount of food consumed.
Weigh the portion sizes of salted replicas of the actual foods consumed by the respondent.

A major advantage of a 24h recall is that the respondent burden is small so that compliance is generally high. In addition, the method is quick and relatively inexpensive, and can be used equally well with both literate and illiterate respondents. Nevertheless, any 24h interview protocol must be standardized, pretested,and then piloted prior to use. Standardization is particularly important in large-scale national surveys and for comparisons across countries(Slimani et al., 2000; Vossenaar et al., 2020). Adherence to the interview protocol and accuracy of food coding by the interviewers should be checked periodically during the survey, and the interviewers must be retrained if required to minimize interviewer bias (Chapter 5). Detailed suggestions on how to conduct the interview can be found in Hughes (1986) who stressed that leading questions and judgmental comments should be avoided. An indirect approach employing open-ended questions is recommended. This enables respondents to freely express their feelings so that answers are not biased. Piloting should be undertaken in an area near the study site, using respondents similar to those who will participate in the actual study.

When 24h recalls are used to characterize the average usual intake of a population group, the respondents should be representative of the population under study. In addition, the survey should be conducted in such a way that all days of the week are equally represented. In this way, any day-of-the-week effects on food or nutrient intakes will be taken into account (Chapter 6). Seasonality must also be considered as it affects food availability.

Guidelines on how to plan, design, and conduct large-scale 24h recall dietary surveys in low-and middle-income countries have

been developed by the Intake Center for Dietary Assessment (Vossenaar et al., 2020; Deitchler et al., 2020), An outline of the main tasks that must be completed prior to the conduct of a large-scale 24h recall survey is presented in Figure 3.1. Intake.org also provide recommended specifications for dietary scales and procedures to test their accuracy and precision (Vossenaar et al., 2020).

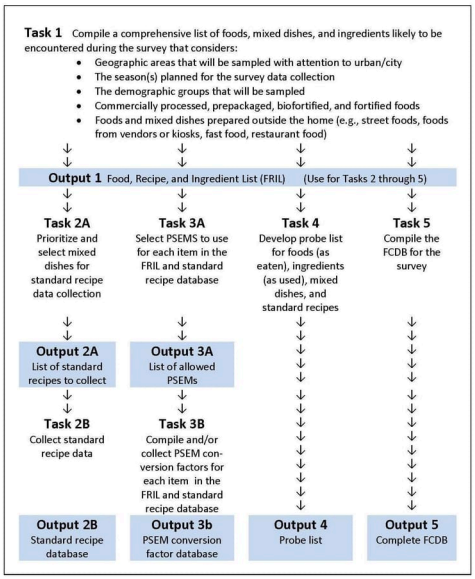


Figure 3.1. PSEM: Portion size estimation method. FCDB: Food composition database. Guidelines on planning for a large-scale 24-hour recall dietary survey in low- and middle-income countries. From: Vossenaar et al. (2020).

A 24h recall has been used in several national nutrition surveys in low-, middle-, and high-income countries. Examples include the Cameroon (Engle-Stone et al., 2014), Nigeria (Maziya-Dixon et al., 2004), New Zealand (Hennigar et al., 2018). (MOH, 2011), and the US NHANES III survey (Hennigar et al., 2018). Intake.org has provided technical assistance to support national or large-scale dietary surveys in Ethiopia, Jordan, Kenya, Niger, Nigeria, Senegal, Viet Nam, and Zambia. The AMPR method, used in the US NHANES III, has been adapted for use by both the Canadian and Australian national health surveys.

The US National Cancer Institute have developed an Automated Self-Administered 24h recall (ASA24) based on the US AMPR method which is suitable for large-scale epidemiological studies (Subar et al., 2012). The ASA24 web-based system was tested for use with children and adults in Canada (Kirkpatrick et al., 2017). Overall participants were receptive to completing ASA24, although the ASA24 interface and the steps in completing the recalls were not necessarily intuitive, especially for younger children and elderly adults. The investigators emphasized the importance of piloting protocols using online tools and the potential need for tailored resources to support certain age groups.

In general, recall interviews can be conducted on children aged > 12y (Deitchler et al., 2020), and on most adults, except for persons with poor memories (e.g., some elderly). Parents are typically relied upon as proxy reporters for capturing dietary intake for children under five years of age. Children aged from 5–12y should be interviewed along with their primary caretaker, usually the mother, an approach termed assisted reporting. It may be necessary to interview several people if the children are at school, day-care, or play in the homes of friends, to ensure that all foods eaten away from home are reported. Recipe data for school or day-care meals may be required to supplement the food intake data collected in the home. In a study of Canadian preschool children 2–5y in which parents used the online self-administered 24h recall (ASA24), they were able to report the food and beverages that their child had consumed with reasonable accuracy but the accuracy of portion size estimates was low(Wallace et al., 2018).

In the end, the success of the 24h recall, irrespective of the method applied, depends on the respondent's memory, their ability to convey accurate estimates of portion sizes consumed, the degree of motivation of the respondent, and the persistence of the interviewer (Acheson et al., 1980).

3.1.2 Repeated 24h recalls

Twenty-four-hour recalls can be repeated to estimate the average food intake of individuals over a longer time period (i.e., usual

food intake). The number of 24h recalls required to estimate the usual nutrient intake of individuals depends on the day-to-day variation in food intake within one individual (i.e., within-person variation). In turn, this variation is affected by the nutrient under study, the study population, and seasonal variations in intake.

Repeated 24h recalls were recommended as part of a system for measuring food consumption patterns in the United States to account for within-person variability as early as 1981 (NRC, 1981). In the US NHANES 2011–2014, food intakes on two non-consecutive days were recorded, on the first day through using computer-assisted personal recall interviews (CAPI), and on the second day via a computer-assisted systematic telephone interview (CATI), using the AMPR on both occasions (Amoutzopoulos et al., 2018). The second recall interviews were not conducted on all respondents, but instead were repeated on a randomly selected subsample of the population. Tooze (2020) recommends that the repeated 24h recalls should be performed on non-consecutive days 3–10d apart, and when only a random subsample of the population is used, repeated on at least 50 individuals per stratum. Moreover, when episodically consumed foods, food groups or nutrients are of particular interest, then more replicates per person should be collected rather than increasing the sample size from which two replicates are obtained (Tooze, 2020).

Repeated 24h recalls assisted by interviewers were also used for the Dutch National Food Consumption Survey (DNFCS) (van Rossum et al. 2016)and the French Nutrition and Health Survey (Castetebon et al., 2009). Several research groups in the UK have incorporated new technologies to collect 24h recalls, including the two web-based methods, myfood24 (Carter et al., 2015) and INTAKE24 (Simpson et al., 2017); see Section 3.2.3 for more details.

3.1.3 Estimated food records

For the estimated food record, also referred to as a food diary by some researchers, the respondent is asked to record in household measures, at the time of consumption, all foods and beverages eaten (including snacks), for a specified time period. Detailed descriptions of all foods and beverages (including brand names) and their method of preparation and cooking should also be recorded. For mixed dishes such as Spaghetti Bolognese, the amount of each raw ingredient used in the recipe, the final weight of the mixed dish, and the amount consumed by the respondent should be recorded, wherever possible. The information is recorded on a form similar to that shown in Appendix 3-1 , except that household measures are used for food amounts. Usually, the respondent, parent, or caretaker completes the food record, although in low or middle-income countries a local field investigator may perform this task (Dufour et al., 1999). Reactivity, defined as a change in behavior due to awareness that behavior is being or will be measured, may be an issue with this method. The respondent has the opportunity to modify his or her “usual” diet, potentially in a more socially desirable manner or to simplify the recording task (Thompson et al., 2015) (Chapter 5).

Food portion sizes can be estimated by the respondent in a variety of ways. Standard household measuring cups and spoons should be used if possible, supplemented by measurements with a ruler (for meat and cake) and counts (for eggs and bread slices). Unfortunately, errors may arise because the respondent may fail to quantify portion sizes correctly. Additional errors may also arise during the conversion of volumes to weights (Chapter 5), although this latter step is usually completed by the investigator. Details on how to convert portion sizes to weight equivalents are given in Gibson and Ferguson (2008) and Vossenaar et al. (2022).

The number of days included in an estimated record varies, depending on the study objective. When the objective is to obtain an average intake for a group, then only one day per person is required, provided all days of the week are equally represented in the final sample. However, when estimates of usual intakes of each person are required, then the number, selection, and spacing of the days required per person depends on the factors described for the repeated 24h recall (Section 3.1.2). Weekend days should always be proportionately included in the dietary survey period for each person, to account for potential day-of-the-week effects on food and nutrient intakes. This problem is discussed in more detail in Chapter 6.

The European Prospective Investigation of Cancer (EPIC) study in Norfolk, U.K. collected food and nutrient intakes from 2117 men and women using a 7d estimated food diary. EPIC was a large multicenter prospective study aimed at investigating the relationship between nutrition and various life-style factors and the etiology of cancer and other chronic diseases. The study involved 23 regional centers located in ten countries, and involved a total cohort of about 480,000 subjects. The respondents of the EPIC study in Norfolk were provided with a “diet diary” — a 45-page colored booklet in which they were asked to record the description, preparation, and amounts of foods eaten over seven consecutive days. Food portion sizes were estimated by the respondents in terms of household measures, with the help of 17 sets of color photographs of small, medium, and large portions of the different foods. See Vossenaar et al. (2020) for guidance on the development of food photographs for portion size estimation. Details on the EPIC study in Norfolk Details are given in (Bingham et al., 2001). The Danish National Survey of Diet and Physical Activity (4–75y) also used a seven-day estimated food diary with a paper format (Pedersen et al., 2015).

A four-day estimated food diary, recorded using a paper form, is used by the U.K. National Diet and Nutrition Survey rolling

program. This is a continuous cross-sectional survey designed to collect detailed information on food and nutrient intakes and nutritional status for a range of ages and across the social strata in the UK. The field work for this rolling program began in 2008 and is on-going with about 1000 people surveyed per year (Ziauddeen et al., 2018).

3.1.4 Weighed food records

Weighed food records are more frequently used in the United Kingdom and Europe because weighing scales are often used for food preparation in these regions. In the earlier British National Diet and Nutrition Surveys of adults and children (Ashwell et al., 2006), seven-day weighed food records were used. However, when the rolling program was introduced in 2008, the seven-day weighed food records were replaced by a four-day estimated food diary due to concerns about respondent burden.

A weighed food record is the most precise method available for estimating usual food and nutrient intakes of individuals. It is the preferred method when diet counseling or correlation of intakes with biological parameters are involved. In a weighed record, the subject, parent, or caretaker is instructed to weigh all foods and beverages consumed by the subject during a specified time period. Details of methods of food preparation, description of foods, and brand names (if known) should also be recorded. For mixed dishes such as Spaghetti Bolognese, the weight of the portion consumed should be recorded, along with the weights and description of all the raw ingredients, including flavors and spices used in the recipe, as well as the final total weight of the mixed dish. The method of recording is similar to that shown for a 24h recall (Table 3.1), with the weight of the food items being recorded under "Amount."

If occasional meals are eaten away from home, respondents are generally requested to record descriptions of the amounts of food eaten. The nutritionist can then buy and weigh a duplicate portion of each recorded food item, where possible, to assess the probable weight consumed. Alternatively, if appropriate, the nutritionist can telephone a restaurant to obtain details of the portion sizes consumed.

As with the estimated record, the number, spacing, and selection of days necessary to characterize the usual nutrient intakes of an individual using the weighed record depend on the within-person variation in food intake, which, in turn, depends on the nutrient of interest, the study population, and any seasonal variation of intake. Again, week-end days should be proportionately included to account for any weekend effect on the nutrient intake. If a weighed food record method is to be used, respondents must be motivated, numerate, and literate. Reactivity may be an issue with this method, and interfere with the assessment of usual dietary intake. Respondents may change their usual eating pattern to simplify the weighing process or, alternatively, to impress the investigator (Cameron and van Staveren, 1988; Thompson et al., 2015) (Chapter 7). In addition, respondent burden for a weighed food record is higher than for an estimated record or for a 24h recall, so individuals may be less willing to cooperate. Reproducibility, however, is greater in the weighed record than in the estimated record method because the portion sizes are weighed, although significant underreporting (Chapter 5) may still occur.

3.1.5 Dietary history

The dietary history method (Burke, 1947) attempts to estimate the usual food intake and meal pattern of individuals over a relatively long period of time — often a month. This interview method was originally designed to be carried out by a nutritionist trained in interviewing techniques. More recently, computerized versions have been developed which provide standardized methods for data collection and probing, and minimize potential interviewer bias in responses (Kohlmeier et al., 1997).

Initially, the dietary history had three components. The first component was an interview about the usual overall eating pattern of the subject, both at mealtimes and between meals. Such information included detailed descriptions of foods, their frequency of consumption, and usual portion sizes in common household measures. “What do you usually eat for breakfast?” is a typical question that might have been included in the interview.

The second component served as a crosscheck and consisted of a questionnaire on the frequency of consumption of specific food items. This part was used to verify and clarify the information on the kinds and amounts of foods given as the usual intake in the first component. Questions asked related to specific foods, such as: "Do you like or dislike milk." A 24h recall of actual intake may also have been included at this stage.

In the third component, subjects recorded their food intake at home for three days. Portion sizes at this stage were estimated using a variety of techniques, including standard measuring cups and spoons, common utensils, commercial plastic food models, photographs, or real foods. Today, the original dietary history method is seldom used in this three-part format, the third component being commonly omitted.

The time periods covered by the dietary history method vary. The maximum time period that can be used has not been definitely established. When shorter time frames (i.e., one month) are used, reproducibility and validity are apparently higher than for longer periods (see Chapter 7). Measurements of food intake over one-year periods are probably unrealistic unless seasonal variations in food intakes are taken into account.

Dutch investigators used a three-part dietary history method covering one month to record usual food consumption on weekdays, Saturdays, and Sundays separately (van Staveren et al., 1985). This approach takes into account the potential effect of weekends on nutrient intake. The portion size of foods most frequently consumed in this study were weighed by a dietitian in the home. A weighted daily average intake was then calculated from the data, using the following formula:

$$((5 \times \text{ Weekday }) + \text{ Saturday } + \text{ Sunday })/7$$

A modified version of this dietary history method was adopted in the Survey in Europe on Nutrition in the Elderly: A Concerted Action (SENECA). This multicenter survey was designed to examine cross-cultural variations in the nutrition, lifestyle, health, and performance of elderly Europeans (Euronut - SENECA, 1991). The method involved the completion of a three-day estimated record, followed by an interview during which respondents were questioned about their usual dietary intake over the past month. Portions of the most commonly eaten foods were weighed by the interviewer (van Staveren et al., 1996).

The recording of a dietary history can be very labor intensive, with interviews taking up to 2h per respondent (Slattery et al., 2000). Several investigators have reported that the dietary history tends to overestimate nutrient intakes, when compared with results from weighed records. Nes et al.(1991), for example, used the dietary history developed for the SENECA study and showed that the method generated consistently higher intakes of energy and nutrients than three-day weighed records. Livingstone and Robson (2000) reported similar findings in a study of children and adolescents, but claimed that the results obtained from the dietary history were more representative of habitual intake than those obtained from seven-day weighed records. In general, because dietary histories, unlike food frequency questionnaires, do not limit the variability in the responses, they overcome many of the limitations of a food frequency questionnaire. However, the absence of a standardized format for the dietary history method limits comparison of its usefulness across studies (Thompson et al., 2015).

3.1.6 Food frequency questionnaire

The food frequency questionnaire, sometimes referred to as a diet history questionnaire, aims to assess the usual frequency with which food items or food groups are consumed during a specified time period. It was originally designed to provide descriptive qualitative information about usual food-consumption patterns. With the addition of portion-size estimates and the introduction of improved computerized self-administered questionnaires, the method has become semi-quantitative, allowing the derivation of energy and selected nutrient intakes (Willett et al., 1985; Block et al., 1986).

In its simplest form, the questionnaire consists of a list of foods and an associated set of frequency-of-use response categories (Appendix 3-2). The list of foods may focus on specific groups of foods, particular foods, or foods consumed periodically in association with special events or seasons. Alternatively, the food list may be extensive to enable estimates of total food intake and dietary diversity to be made. The frequency-of-use response categories may be daily, weekly, monthly, or yearly, depending on the study objective.

Specific combinations of foods can be used as predictors for intakes of certain nutrients or non-nutrients, provided that the dietary components are concentrated in a relatively small number of foods or specific food groups. Examples include the frequency of consumption of fresh fruits and fruit juices as predictors of vitamin C intake (Tsugane et al., 1998), green leafy vegetables and carrots (O'Neill et al., 2001), or fruit and vegetable intake as predictors of carotenoid intakes (Whitton et al., 2017), whole grain cereals, legumes, nuts, fruits, and vegetables as predictors of dietary fiber intakes (Merchant et al., 2003), and dairy products as predictors of calcium intakes (Barr et al., 2001; Horiuchi et al., 2019). The method can also be used to assess the intake of fats and cholesterol (Feunekes et al., 1993; Eng and Moy, 2011; Riordan et al., 2018), artificial sweeteners (Dewinter et al., 2016), certain contaminants present in specific foods (MacIntosh et al., 1997; Filippini et al., 2018), alcohol (Bazal et al., 2019), and condiments (Leyvraz et al., 2018).

The food frequency questionnaires should feature simple, well-defined foods and food categories. Open-ended questions should be avoided as preformatted lists of food categories act as a memory prompt. The method may use a standardized interview, a self-administered machine-readable printed questionnaire, or a computer-administered questionnaire. Most questionnaires usually list 80 to 120 individual items and take from 15–30minutes to complete (see abbreviated example given in Appendix 3.2). Hence, the food frequency questionnaire imposes less burden on respondents than most of the other dietary assessment methods. The results are easy to collect and process and are generally taken to represent usual intakes over an extended period of

time, and hence are not affected by day-to-day variability. As food frequency questionnaires are designed to provide retrospective information about diet, they are often used in retrospective case-control studies (Schink et al., 2019). However, their validity and feasibility for estimating food intakes in the remote past has not always been clearly established (van Staveren et al., 1986; Dwyer and Coleman, 1997; Ambrosini et al., 2003; Barrett et al., 2019).

WHO has issued guidelines on a Healthy Diet (WHO, 2020). A (Healthy Eating Index) (HEI-2015) has been developed by the USDA, that measures alignment of US diets with the 2015-2020 Dietary Guidelines for Americans. The HEI-2015 has 13 dietary components in total, of which nine are "adequacy" components (those recommended for inclusion in a healthy diet) and four "moderation" components that should be consumed sparingly. These dietary components are listed in Table 3.3. This table also shows the scoring standard for each of the dietary components. The scores for each of the 13 components derived from the individual's intake, are added to give an HEI Index score (maximum = 100) for the individual.

Table 3.3 HEI–2015 Components & Scoring Standards. From HEI (2020).

[1]Intakes between the minimum and maximum standards are scored proportionately.

[2]Includes 100% fruit juice.

[3]Includes all forms except juice.

[4]Includes legumes (beans and peas). *[5]Includes all milk products, such as fluid milk, yogurt, and cheese, and fortified soy beverages.*

[6]Includes legumes (beans and peas).

[6,7]Includes seafood, nuts, seeds, soy products (other than beverages), and legumes (beans and peas).

[8]Ratio of poly- and monounsaturated fatty acids (PUFAs and MUFAs) to saturated fatty acids (SFAs).

HEI–2015[1] Components & Scoring Standards			
Component	**Max. points**	**Standard for max. score (per 1,000 kcal)**	**Standard for min. score (0)**
Adequacy			
Total Fruits[2]	5	≥ 0.8 cup equiv.	No Fruit
Whole Fruits[3]	5	≥ 0.4 cup equiv.	No Whole Fruit
Total Vegetables[4]	5	≥ 1.1 cup equiv.	No Vegetables
Greens and Beans[4]	5	≥ 0.2 cup equiv.	No Dark Green Veg. or Legumes
Whole Grains	10	≥ 1.5 oz equiv.	No Whole Grains
Dairy[5]	10	≥ 1.3 cup equiv.	No Dairy
Total Protein Foods[6]	5	≥ 2.5 oz equiv.	No Protein Foods
Seafood and Plant Proteins[6,7]	5	≥ 0.8 oz equiv.	No Seafood or Plant Proteins
Fatty Acids[8]	10	(PUFAs + MUFAs) /SFAs ≥ 2.5	(PUFAs + MUFAs) /SFAs ≤ 1.2
Moderation			
Refined Grains	10	≤ 1.8 oz equiv.	≥ 4.3 oz equiv.
Sodium	10	≤ 1.1 gram	≥ 2.0 grams
Added Sugars	10	≤ 6.5% of energy	≥ 26% of energy
Saturated Fats	10	≤ 8% of energy	≥ 16% of energy

Kant et al. (2000) used food frequency information to calculate a "Recommended Food Score" (RFS) to evaluate the consumption

of foods considered to be consistent with the U.S dietary guidelines existing at that time. In a prospective study of diet quality and mortality in women from the United States, they showed that the RFS was inversely associated with all-cause mortality. Since that time, the RFS has been used in Australia as a diet quality index for preschoolers (Burrows et al., 2014), and in Korea to investigate links with physical performance among the elderly (Jeong et al., 2019), and depression, anxiety, and quality of life (Lee et al., 2019).

Many recent users of food frequency questionnaires have quantified portion sizes of food items of interest, often using photographs (Nelson et al., 1994; Amougou et al., 2016). Portion sizes can be ranked as small, medium, and large, preferably based on age and sex-specific portion size data generated from country-specific national nutrition surveys (Willet et al., 1985; Block et al., 1986; Bohlscheid-Thomas et al., 1997). Note that inclusion of information on portion sizes produces semi-quantitative food frequency data (Appendix 3-3). This can be converted to data on energy and nutrient intakes by multiplying the fractional portion size of each food consumed per day by its energy and nutrient content, obtained from appropriate food composition data. The results are then summed to obtain an estimate of an individual's total daily energy and nutrient intake.

Block et al. (1986) derived a food list for a food frequency questionnaire with portion sizes from the NHANES II results. Food items selected were based on the frequency of consumption of certain specific food items and which contributed significantly to the total population intake of energy and each of 17 nutrients. Serving sizes were estimated from observed portion size distributions in the NHANES II data. Medium serving sizes for each food were specified in the food frequency questionnaire, and the respondent indicated whether his or her usual serving size was small, medium, or large, as shown in Appendix 3.3. A specialized food composition database was developed for use with this food frequency questionnaire. A very similar approach has been used to design semiquantitative multi-ethnic food frequency questionnaires (Deurenberg-Yap et al., 2000; Beukers et al., 2015).

An improvement to the Block food frequency questionnaire termed the Diet History Questionnaire (DHQ) has been developed. The DHQ consists of 124 food items, portion sizes, and dietary supplement questions and takes one hour to complete. Results of a comparative validity study showed that the performance of the DHQ is better than that of the food frequency questionnaires of both Block and Willett(Subar et al., 2001).

The semi-quantitative food frequency questionnaire has become a widely used tool in dietary assessment. Country-specific semiquantitative food frequency questionnaires containing between 130 and 300 food items were used in the EPIC study to estimate individual usual food intakes (Margetts and Pietinen, 1997). In the EPIC study in Norfolk, U.K. (Bingham et al., 2001) respondents (n = 23,003) estimated how frequently foods were eaten over the past year, from nine possible frequency-of-use response categories from a list of 130 foods. Reduced versions containing only 60 food items that require only 17 min to administer by an interviewer are available; even the full 98-item Block questionnaire requires only 30-35 min of interviewer time (Block et al., 1990). In some countries, a semi-quantitative food frequency questionnaire has been used in national nutrition surveys (e.g., 1995 Australia National Dietary Survey) (Williams, 2005).

More recently in Australia, an on-line suite of validated semi-quantitative food frequency questionnaires, the Australian Eating Surveys (AES) have been developed. The AES is a 120-item food frequency questionnaire with fifteen supplementary questions, including food and sedentary behaviors and supplements, with the aim of capturing the usual dietary intakes of children, adolescents, and adults over the previous six months. The AES takes only 15-20 min to complete on-line and can generate a personalized dietary feedback in real time as an incentive to encourage participation and enhance the response rate (Collins et al., 2014).

In many low-income countries no national food consumption surveys have been conducted so the key information needed to develop a food frequency questionnaire is often not available. Such information includes a listing of foods commonly consumed in the study population, details on the way the foods are typically prepared or consumed, and the usual range of portion sizes consumed. In response to this need, Hotz and Abdelrahman (2019) have developed some simple methods to obtain food listing and portion size distribution estimates for use in semi-quantitative food frequency questionnaires.

Table 3.4 Ref = reference quintile. Multivariate model: age at baseline, total calorie intake, smoking history, cancer (yes/no) hypertension diagnosis (yes/no), use of antidepressants, elevated cholesterol, physical activity level, body mass index, cardiovascular disease, multivitamin use, intake of alcohol, total calorie intake, profession, missing indicator for SCF, and number of dietary assessments during1986–2002. From Yuan et al. (2019).

Odds Ratios (95% CIs) for poor Subjective Cognitive Function (SCF), compared with good function, associated with total vegetable, fruit, and fruit juice intakes					
Quintile of intake (n = 27,842 men)	Q1	Q2	Q3	Q4	Q5
Total vegetable intake Median, servings/d	1.7	2.5	3.2	4.1	5.7
Odds Ratio 95% Confidence Interval	Ref	0.92 (0.79, 1.08)	0.85 (0.73, 0.99)	0.71 (0.60, 0.83)	0.62 (0.52, 0.74)
Total fruit intake Median, servings/d	0.5	1.1	1.5	2.0	3.1
Odds Ratio 95% Confidence Interval	Ref	1.00 (0.85, 1.18)	0.96 (0.81, 1.13)	0.81 (0.68, 0.96)	0.79 (0.66, 0.94)
Total fruit juice intake Median, servings/d	0.1	0.4	0.7	1.0	1.5
Odds Ratio 95% Confidence Interval	Ref	0.78 (0.67, 0.91)	0.81 (0.70, 0.94)	0.66 (0.57, 0.77)	0.65 (0.55, 0.76)

Food frequency questionnaires are often used by epidemiologists studying associations between dietary habits and disease. For example Yuan et al. (2019) in their study of the relationship between long-term intake of vegetables and fruits and subjective cognitive function in US men, used this approach (Table 3.4). See also: (Willett, 1994; Harris et al., 2018; Zhong et al., 2019). In such studies, the food frequency questionnaires must be semi-quantitative, with the ability to rank subjects on the basis of their intakes, so that subjects with low intakes can be separated from those with high intakes. This permits the calculation of the odds ratio or relative risk of disease in relation to intake of certain foods, food groups, or nutrients (Masson et al., 2003).

Box 3.1 summarizes the five methods discussed above that can be used to assess the food consumption of individuals and their uses and limitations. For a more detailed review of the strengths and weaknesses of each of these methods for use in low- and middle-income countries, see (FAO, 2018).

Box 3.1 Uses and limitations of methods used to assess the food consumption of individuals.

24h recall. Subject or caretaker recalls food intake of previous 24h in an interview. Quantities estimated in household measures using food models as memory aids or to assist in quantifying portion sizes. Nutrient intakes calculated using food composition data.

- **Uses and Limitations** Useful for assessing average usual intakes of a large population, provided that the sample is truly representative and that the days of the week are adequately represented. Used for international comparisons of relationships of nutrient intakes to health and susceptibility to chronic disease. Inexpensive, easy, quick, with low respondent burden so that compliance is high. Large coverage possible; can be used with illiterate individuals. Element of surprise so less likely to modify eating pattern. Single 24h recalls likely to omit foods consumed infrequently. Relies on memory and hence unsatisfactory for the elderly and young children. Multiple replicate 24h recalls can be used to estimate the usual intakes of individuals.

Estimated food record. Record of all food and beverages “as eaten” (including snacks), over periods from one to seven days. Quantities estimated in household measures. Nutrient intakes calculated using food composition data.

- **Uses and Limitations** Used to assess actual or usual intakes of individuals, depending on number of measurement days. Data on usual intakes used for diet counseling and statistical analysis involving correlation and regression. Accuracy depends on the conscientiousness of subject and ability to estimate quantities. Reactivity may be an issue. Longer time frames result in a higher respondent burden and lower cooperation. Subjects must be literate unless a trained researcher conducts the estimated food record in the home.

Weighed food record. All food consumed over a defined period is weighed by the subject, caretaker, or assistant. Food samples may be saved individually, or as a composite, for nutrient analysis. Alternatively, nutrient intakes calculated using food composition data.

- **Uses and Limitations** Used to assess actual or usual intakes of individuals, depending on the number of measurement days. Accurate but time consuming. Setting must permit weighing. Subjects may change their usual eating pattern to simplify weighing or to impress investigator — termed — reactivity. Requires motivated, and willing participants. Expensive. For illiterate participants, weighed food records can be conducted by a trained research assistant in the home.

Dietary history. Interview method consisting of a 24h recall of actual intake, plus information on overall usual eating pattern, followed by a food frequency questionnaire to verify and clarify initial data. Usual portion sizes recorded in household measures. Nutrient intakes calculated using food composition data.

- **Uses and Limitations.** Used to describe usual food or nutrient intakes over a relatively long time period, which can be used to estimate prevalence of inadequate intakes. Such information is used for national food policy development, for food fortification planning, and to identify food patterns associated with inadequate intakes. Labor-intensive, time-consuming, and results depend on skill of interviewer. The absence of a standardized format limits comparison of its usefulness across studies.

Food frequency questionnaire. Uses comprehensive or specific food item list to record intakes over a given period (day, week, month, year). Record is obtained by interview or self-administered questionnaire. Questionnaire can be semi-quantitative when subjects asked to quantify usual portion sizes of food items, with or without the use of food models.

- **Uses and Limitations** Designed to obtain qualitative, descriptive data on usual intakes of foods, classes of foods, or nutrients (if semi-quantitative) over a long time period. Useful in epidemiological studies for ranking subjects into broad categories of low, medium, and high intakes of specific foods, food components, or nutrients, for comparison with the prevalence or mortality statistics of a specific disease. Can also identify food patterns associated with inadequate intakes of specific nutrients. Method is rapid, with low respondent burden and high response rate, but accuracy is lower than for other methods.

This page titled 3.1: Methods for measuring food consumption of individuals is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

3.2: Technical improvements in food consumption measurements

The increasing evidence of the relationship between diet and chronic disease has led to a number of technical advances in measurements of food consumption for individuals. They include the use of the telephones in surveys, digital photographs of food portions displayed on a computer/tablet or captured with a cell-phone camera, as well as web-based tools, mobile apps, and image-based assessment. These innovative technologies aim to improve the speed, accuracy, and participation, and reduce reporting burden, bias, and the cost of collecting and analyzing dietary intake data during large scale epidemiological studies as well as national surveys. A summary of the strengths and limitations of these innovative technologies designed to improve dietary assessment methods is shown in Table 3.5. A robust assessment of the performance, cost, and response rates should be undertaken, however, prior to the adoption of any new-technology-based method.

Table 3.5 Strengths and limitations of innovative technologies to improve dietary assessment methods. From: Dietary Assessment: A resource guide. FAO (2018).

All innovative technologies used in dietary assessment	
Strengths	**Limitations**
Costs for data collection can be lower (less need for person-to-person interaction)	Larger up-front investments (i.e. purchase of mobile phones, cameras, computers, software development, etc.)
Convenient for users, good acceptability, may improve compliance	Risk of losing devices
Do not rely on respondent's memory	Risk of technical problems (i.e. low battery, loss of Internet connection) could impede data collection
Record of qualitative information (e.g. date and time of recording)	A backup method is required to collect information, if technical problems occur
Significantly cuts down data processing time	
Personal Digital Assistant (PDA)	
Strengths	**Limitations**
Is portable and can be easily carried by study participants	Face-to-face training of the participants is required
Facilitates real-time data collection, entry and coding	Low level of dietary data details because of pre-coded food listings
It is possible to set an alarm within the PDA to alert participants to record their food intake	Increases the respondent burden compared with pen and paper records, due to the extensive list of foods (depending on each PDA)
Can be programmed to allow participants strict access to the dietary intake software only	Reports of difficulty using the search function and inability to find certain foods
Image-assisted methods i.e. digital photographs	
Strengths	**Limitations**
Easy to use	Participants may forget to take some images

All innovative technologies used in dietary assessment	
Strengths	**Limitations**
Suitable for low literacy populations (if the technologies are easy-to-use, e.g. digital cameras)	Not all information can be captured with a single photograph/image
Quality of digital cameras keeps improving and pictures with higher resolutions can help improve the accuracy of analysis	Difficulties in estimating portion size of food consumed from common mixed dishes
Suitable for subjects with memory impairments and for children	Lack of details about cooking methods
Lower under-reporting compared with some traditional methods	Probably more limited accuracy for countries with a wide range of mixed dishes (e.g. Asia) Still needs a written record of foods obscured in photos, and when details of ingredients are required
Image-assisted methods i.e. mobile-based technologies	
Strengths	**Limitations**
Possible higher quality control of data because of shorter delays and real-time responses	It is costly and time-consuming to develop the application as an interface, and the software for automated portion size estimation
Possibility of sending reminders	Certain types of foods, such as mixed dishes (soups, stews, casseroles, etc.) can be difficult to analyse with automated image analysis
Internet access would allow respondents to send instant photos, thus minimizing systematic mistakes	Requires certain level of literacy Requires network/Internet access (for real-time data collection)
Interactive computer and web-based technologies	
Strengths	**Limitations**
Efficiency in terms of recording information and data processing (i.e. reduced costs and interviewer workload)	Some imaging algorithms presently fail to identify foods correctly and to accurately estimate the quantity of food in the computerized images
Increased levels of quality control	Need adaptation of the software to local settings
Include interactive visual and audible aids	Require high levels of literacy and computer skills from the participants
Suitable for large, geographically dispersed samples; can include different countries/languages (web-based)	Possibility of collecting less food details (i.e. ingredients, methods of preparation, etc.)
Data processing can be finalized at any time and location (web-based) Can provide personalized dietary feedback (web-based) Possibility of sending interactive reminders (web-based)	Require Internet access (web-based)

3.2.1 Telephone

A telephone survey that is well-designed and carefully-administered appears to be a promising method of obtaining dietary information. In particular, telephoned 24h recalls are being increasingly used. The USDA has conducted several studies to examine the feasibility of using telephone follow-up surveys instead of mail follow-up for 24h recalls. Results have been promising: response rates for telephone follow-up were much greater than those for mail follow-up. Casey et al. (1999) carried out a validation study in which the results of 24h recalls conducted over the telephone were compared with in-person recalls collected in the 1994-1996 USDA Continuing Survey of Food Intakes by Individuals; strong correlations were reported. As a result, since 2002 NHANES have used the telephone to administer the second recall day 3–10d later using the AMPR method (Hennigar et al., 2018).

Several other large-scale telephone dietary surveys have been used in the United States, some of which have used a food frequency instrument rather than a 24h recall. Lyu et al.(1998) successfully showed that agreement between telephone and face-to-face interviews of a semi-quantitative food frequency questionnaire in Hawaii made up of 115 food items was good and unaffected by age, gender, ethnicity, or education of the respondents. These investigators did recommend mailing photographs of foods in three portion sizes in advance, to help respondents estimate amounts eaten more accurately.

Many of the smaller telephone dietary studies have been conducted on adult women (Casey et al. 1999; Tran et al., 2000; Yanek et al., 2000). Very few have been carried out on adolescents and adult men (Bogle et al., 2001). In college students, food intakes by telephoned recalls have been compared against actual intakes determined surreptitiously in the college cafeteria (Krantzler et al., 1982). More studies are needed, however, among certain life-stage groups, to establish the validity of 24h recalls or food frequency questionnaires administered over the telephone. Under-reporting of self-reported food intakes may still occur in telephone recalls, as they do with in-person recalls; this was observed, for example, when total energy intakes derived from telephone-administered 24h recalls were compared with total energy expenditure measured by doubly labeled water (Tran et al., 2000) (see Chapter 7).

Telephone surveys do have several advantages. These include their ability to reach a large number of persons at perhaps less than half the cost of face-to-face surveys. As well, with the advent of computer-assisted telephone interviewing, the interviews can be readily standardized, queries can be clarified, and responses can be coded immediately, during an interviewing time that is much shorter than a face-to-face interview. As a result, the response rate is enhanced.

Potential interviewer biases can be eliminated by using computer-assisted telephone interviewing, an approach used with the AMPR in the NHANES surveys. However, other sources of bias may occur with telephone surveys; these may arise from non-coverage and nonresponse. In the United States, although over 87% of the population owns a telephone, subgroups such as the poor, certain minorities, and the` elderly still have fewer telephones than the general population. Such differential coverage can introduce bias in national surveys unless alternative compensating strategies are employed. Strategies may include using supplementary face-to-face or mail interviews for persons without telephones; statistical adjustments employing weighting for sex, age, race, and income; selective over-sampling of the non-telephone users; and random-digit dialing to contact those with unlisted numbers. Nonresponse is also a source of potential bias that influences telephone surveys. This is especially a problem among certain subgroups such as the ethnic minorities for whom language barriers may be considerable, the elderly who may have hearing difficulties, and persons with less education.

Despite these limitations, telephone dietary surveys are a practical, economical, and valid alternative to the conventional face-to-face methods for large-scale epidemiological studies and nutrition surveys and for developing and evaluating community-based nutrition interventions in higher income countries.

3.2.2 Portion size estimation aids

Errors in quantifying the portions of food consumed are often the largest error in most food consumption surveys, unless weighed food records are used. As a result, several types of portion-size measurement aids (PSMAs) have been developed in an effort to enhance the accuracy of portion size estimates (see Chapter 5 for more details). PSMAs can be divided into two categories: (1) non-photographic aids and (2) photographic aids. Canada was one of the first countries to use non-photographic aids in its National Nutrition Survey in 1973. The aids consisted of a collection of three-dimensional graduated food models of various volumes and surface areas prepared from paper-mache, wood, or hardboard. A range of graduated food models was used to prevent the tendency to generate a “direct” response. The latter phenomenon is observed when plastic food models representing only “average” portion sizes are used (Samuelson, 1970). The Canadian surface-area models were accompanied by standard thickness indicators made of hardboard squares. These were used during the 24h recalls to assist in assessing the overall size and thickness of foods such as cheese, cold meats, cakes and cookies. Use of thickness indicators is critical for assessing portion

sizes of intact cuts of meat, especially when irregular in shape.

Since 1973, graduated food models (Figure 3.2) have been used with 24h recalls in many national food consumption surveys, including those in the United States (NHANES, 2017) and New Zealand (MOH, 1997). In the dietary component of the NHANES nationwide surveys, the tool developed to measure portion sizes in the 24h recalls consists of a food model booklet of 32 life-size two-dimensional drawings of household vessels (glasses, mugs, bowls), abstract shapes (mounds and spreads), and geometrical models (circles, a grid, wedges, and thickness bars), together with a set of measuring cups, spoons, and rulers (USDA 2002+); More details are given in Chapter 5. In low-income countries, PSMAs for 24h recalls may be salted replicas of staple foods, real foods, and local household utensils (e.g., bowls, plates, spoons). These are often used in combination with items to simulate liquid and solid foods (e.g., water or raw rice), manipulative models (modelling clay or play dough), tape measures or rulers and graduated food models (Gibson and Ferguson 2008). Guidelines on PSMAs for use with 24h recall methods in low-income countries are available from www.intake.org. The use of less than five different PSMAs are recommended for surveys in low-income countries. (Vossenaar et al., 2020).



Figure 3.2. USDA Food Model Booklet

More recently, photographs are being increasingly used as PSMAs. They may include pictures of individual foods or meals which have been drawn, printed, or displayed digitally, often in color and sometimes shown next to a standard sized object (e.g., a coin) or gridlines to aid the respondent's perception of scale. For surveys using 24h recalls, a series of graduated portion-size photographs (i.e., amount consumed on any one occasion) for each food item are commonly displayed during the interview, whereas for surveys employing food frequency questionnaires, either an average serving size for each food (i.e., average amount served in one helping) or a series of serving sizes (small, average, large) are often used. The graduated portion-size photographs are often bound together in a photographic atlas.

Practical guidelines on how to develop a photographic atlas are given by Nelson and Haraldsdottir (1998) and Vossenaar et al. (2020). Factors that must be considered in relation to the format of the photographs include size of the image, number and range of portions sizes depicted, and the interval between portion sizes. In the EPIC study, for example, there was a 25% difference between portion sizes to allow a real visual perception of differences in size (Vilela et al., 2018). Nelson et al. (1994) used portion weights from the British Adult Dietary Survey ranging from the 5th to the 95th percentile for a series of eight photographs for each food. Other important factors that should be standardized include the order of presentation of the photographs, labels used, angle at which the photograph was taken, background and use of reference objects for scale, color versus black and white, and use of one versus several foods on a plate.

Increasingly digital photos of food portions with multiple images displayed on a computer or a tablet are replacing printed photos to reduce the cost of printing atlases and their transportation to different interview sites. Nichelle et al. (2019) compared the accuracy of three portion sizes (small, medium, large) from printed photographs in the Brazilian GloboDiet atlas with digital photos displayed on a computer screen for 20 foods selected from the Brazilian GloboDiet atlas. The mean error (difference between the estimated and true portions) was not significantly different between the printed and digital photos, with agreement

on using the printed and digital photos being 91% and 90%, respectively.

During the development of the Internet-based, Automated, Self-Administered 24h recall (ASA24), Subar et al. (2012) investigated the best way to provide digital images as portion-size estimation aides. They compared the photographs said to represent portion sizes of foods consumed by the respondents on the previous day with the actual weights obtained by surreptitious weighing. The images were also varied by the angle in which the photographs were taken, the type of image, and the number and size of digital images presented in an effort to determine how best to present the digital images to facilitate accuracy. None of the photographs were labeled with portion-size information. Results indicated that accuracy was not affected by the type of image, whether the images were presented simultaneously vs. sequentially, or the size and number of images presented, although there was a tendency for eight rather than four images to be more accurate. Based on these results, aerial photographs are used for portion size estimation in the ASA24.

Alternatively, real-time images of foods and beverages consumed by the respondent, and captured passively, or actively with a cell-phone camera (Arab et al., 2011), digital camera (Lazarte et al., 2012), or wearable camera (Gemming et al., 2013), can be used as PSMAs during a 24h recall the following day. The active methods require the respondents to take pictures in the appropriate way (e.g., 45°angle, arms distance from the plate, with a size reference included for scale) before and after eating all meals, and to add accompanying notes. However, all of these tasks necessitate training the respondents. As a result, for low-literacy populations, the passive techniques are more suitable, where an individual wears a camera to capture photographs, thus requiring less involvement by the respondent.

In many food consumption surveys in both lower and higher income countries, a mixture of non-photographic and photographic aids have been used as PSMAs. A systematic review that included both non-photographic and photographic PSMAs concluded that culturally appropriate digital or hard-copy photographic PSMAs were more accurate than non-photographic food models and household utensils (Amoutzopoulos et al., 2020). The investigators also emphasized when selecting an appropriate tool for estimating portion sizes, both the setting and population under study must be considered. Most of the studies included in this review were from high income countries. In a study in rural Bolivia, however, the mean energy and nutrient intakes of 45 women (20–52y) obtained using an interviewer-administered atlas of graduated photos of local foods combined with digital food photographs taken by the respondents on the day prior to the 24h recall, were compared with those derived from a weighed record. No significant mean differences were reported in estimated and actual food amounts between the 24h recall using the food photo album and the digital photographs and weighed food records, except for a few foods (rice, potatoes, eggs, vegetables). For the latter foods, differences ranged from –5.4% for vegetables to –6.8% for rice. However, there was a slight underestimation in values between the two methods for some nutrients (Lazarte et al., 2012).

3.2.3 Web-based systems for dietary assessment

Table 3.6 Some of the web-based 24-h recall tools for dietary assessment

Web-based 24-h dietary recall tools for dietary a		
Author	**Tool (Country)**	
(Cade, 2017),	(myfood24), (UK)	Large database with generic and branded
(Bradley et al., 2016),	(INTAKE24), (UK)	Meal-based approach; based on AMPM a
(Lassale et al., 2015),	(France)	Web-based 24h diet record; Food portion
(Thompson et al., 2015),	(ASA24), (USA)	Meal-based approach; uses the five steps optional supplement module
(Arab et al., 2015),	DietDay (USA)	Multiple steps similar to the AMPM appro feedback
(Biltoft-Jensen et al., 2014),	WebDASC (Denmark)	Self-administered tool for children 8–11y; browse or search.
(Vereecken et al., 2014).	CANAA-W (Belgium)	A feasibility study on 131 children and 53

On-line dietary assessment tools based on 24h recalls, food frequency questionnaires, and dietary records are becoming increasingly available. They can be interviewer- or self-administered so data can be collected at a time and location that is

convenient to the respondent. Examples of web-based 24h dietary recall tools are shown in Table 3.6. These include the US National Cancer Institute ASA24 and DietDay (Arab et al., 2011), the UK INTAKE24 (Bradley et al., 2016), and UK myfood24 (Cade, 2017), the French NutriNet Sante (Touvier et al., 2010), WebDAS in Denmark (Biltoft-Jensen, 2014), and the CANAA-W in Belgium (Vereecken et al., 2014). Some of these have been modified for use in other high-income countries such as Canada, Australia and Germany.

The use of automated 24h recall interviews in web-based systems standardizes the interviewing process so there is consistency in the format in relation to the questioning and sequence, food descriptors, and methods used to estimate food portion sizes. Some (e.g., ASA24; INTAKE24; myfood24) use a meal-based approach. Figure 3.3 shows an example of a screenshot from ASA24 for the recall of breakfast. Most are based on a version of the AMPM interview format developed by USDA (Moshfegh et al., 2008).

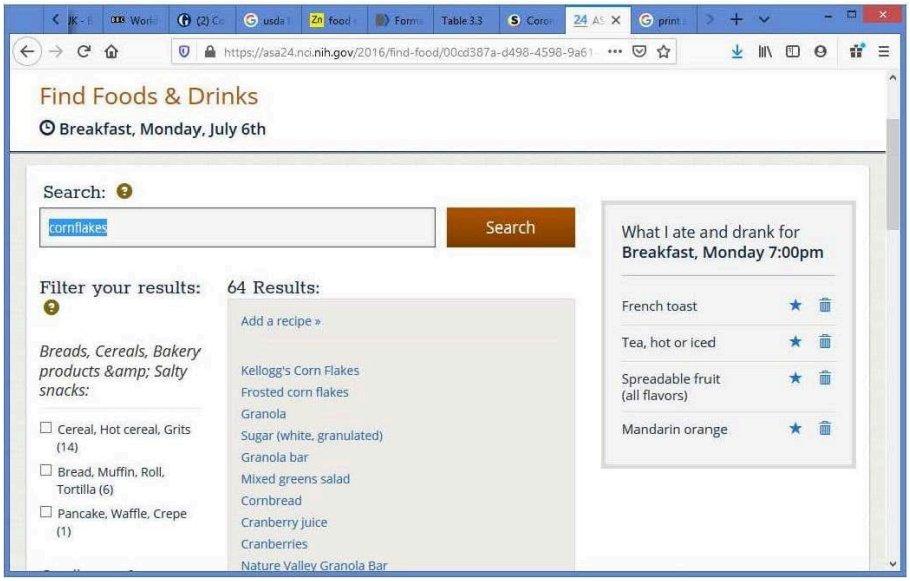


Figure 3.3. A sample screen-view from a web-based assessment tool (ASA24), showing the user selecting, in an unstructured manner, recalled food items for breakfast. Note the user-prompted drop-down list.

Groupings of foods and recipes, as well as the level of detail used to describe foods varies across systems. In most of the applications, ingredients of mixed dishes are entered separately in either the raw or cooked form, with an option to select a generic mixed dish if the ingredients of the recipe are unknown. Some of these on-line systems use only food portion photographs to help respondents report the amount of each food consumed (e.g., ASA24; DietDay), whereas others (e.g., INTAKE24) use multiple options for portion size selection. These may include graduated photographs or images of actual foods, guide photographs for selected prepackaged items (e.g., snacks), with or without additional options such as photographs/pictures of household measures (glasses, cups, bowls, teaspoons etc.) Generally, additional automated prompts for potentially forgotten items are programmed and quality controls are implemented at each step of the interview procedure, with checks made on outlier values. Standardized algorithms are often used to calculate automatically grams of the food consumed, and the amounts of ingredients (either raw or cooked) consumed from a mixed dish. Some of the applications, especially for children, record the amounts served and leftovers rather than the amounts consumed. Most applications have a separate screen for describing and quantifying dietary supplements. Usually, at the end of the interview, a summary is displayed with the opportunity for the respondent to correct or edit information at each stage of the interview. Each web-based system is linked to an in-country food composition database, some of which include branded food items (e.g., ASA24, myfood24), with automated coding for those reported foods and beverages that are in the food composition database. Items not in the database are coded manually.

In low-income countries where reliable internet access may be lacking, especially in rural areas, web-based methods are not feasible for assessing dietary intakes, especially when self-administered, due to the low literacy rates that often exist. To address these concerns in low-income countries, the International Dietary Data Expansion (INDDEX) Project has developed a dietary assessment platform called INDDEX24 (Coates et al., 2017). The platform consists of a mobile application for the collection of off-line individual-level dietary data using the 24h recall method with the AMPM, linked to a web database application so that the users can download the required country-specific food composition data. The dietary recall mobile application uses an open source mobile data collection platform (Commcare) so that it is accessible to researchers at no or low cost. The application is designed so that it is straightforward to use and contains a number of quality control features such as controls to prevent entry

of implausible values, pre-defined portion-size estimation methods depending on the type of food reported, and real time calculation of the respondent's reported energy intake in order to identify possible under- and over-reported intakes (Burrows et al., 2014). The mobile app converts food portions reported in non-standard units into gram-weight equivalents by linking to context-specific portion size conversion factors housed within the web database (Coates et al., 2017). Both the performance and cost of a 24h recall using the INDDEX24 mobile app compared with a pen-and-paper 24h recall have been explored. See Rogers et al. (2022) and Adams et al. (2023) for further details.

3.2.4 Image-based methods of dietary assessment

These are defined as any method in which images are used for identifying foods, or for portion size estimation. Personal digital assistants (i.e., handheld computers) were used initially for these image-based methods, but have now been replaced by camera-enabled cellular phones or smartphones, digital, disposable, or 35mm cameras, wearable cameras, or digital video cameras. Many of these image-enabled devices rely on users manually taking photographs of foods (i.e., active image capture). Alternatively, for those methods termed “passive image capture”, pictures are automatically taken, usually at timed intervals (e.g., every 10 seconds) by the devices (predominantly wearable digital cameras), so that no intervention is required on the part of the user. The majority of tools that use active image capture have been designed for research purposes to enhance the accuracy of self-reported intakes. Fewer passive image capture methods exist and none is fully automated, most relying on specialists to manually estimate portion size (Gemming et al., 2015a).

A few image-based methods are based on a combination of tools. Examples include: SmartPlate (SRI International, 2015), which consists of a camera and sensor-enabled dinner plate, the Diet Data Recorder System (DDSR), a three-part system comprised of a camera-enabled smartphone, laser module, and circuit module (Bell et al., 2017), and the Dietary Intake Monitoring System (DIMS). The latter is a five-part system with a digital camera, weighing scale, infrared thermometer, radio-frequency identification reader (RFID), and a user RFID transponder card (Ofei et al., 2014). Most of the image-based methods are portable. Exceptions are the DIMS which is mounted on a small cart and the SmartPlate which is dinner-plate size and hence less convenient to carry compared to the other smaller devices.

Image-based methods can be further subdivided into three groups: image-based food records, image-assisted food records, and image-assisted 24h recalls. Image-based food records are defined as methods in which images of foods captured during eating episodes serve as the entire basis for the dietary assessment, whereas in image-assisted food records, the images captured are used to enhance or supplement a traditional text-based food record (written or electronic). In image-assisted 24h recalls, the food images captured during eating episodes are used to aid self-reporting during the 24h recall. For more details of image-based methods, the reader is referred to the following reviews (Illner et al., 2012; Gemming et al., 2015a; Boushey et al., 2017).

Several studies have confirmed that image-based methods of dietary assessment are preferred over traditional dietary records, and can enhance self-reported dietary intake by revealing unreported foods and misreporting errors. However, study costs will be increased when respondents have to be provided with a suitable device (Gemming et al., 2015a). With the additional dietary information from the images, there is likely to be an increase in reported energy intake and a decrease in reporting errors, except when the methods are followed incorrectly so the images are of poor quality, or the user forgets to capture images before eating episodes. In addition, details of mixed dishes are difficult to extract from the images, some ingredients may be hidden, and cooking methods omitted. To avoid these problems, respondents should be trained to review the images after taking them, to take a second image if necessary, and to record descriptions of the mixed dishes, including lists of ingredients and quantities. Coding errors associated with image analysis, however, are likely to be less random and less problematic compared to the systematic bias observed when food type and portion size are self-reported.

Figure 3.4 depicts the entire system — the Technology Assisted Dietary Assessment (TADA) system — that commences with the respondent capturing an image with the mobile food record. In this method, respondents use the camera of a mobile devise to capture images of their intake of foods and beverages which they send to a secure server, where trained analysts estimate the reported intakes of energy and nutrients.

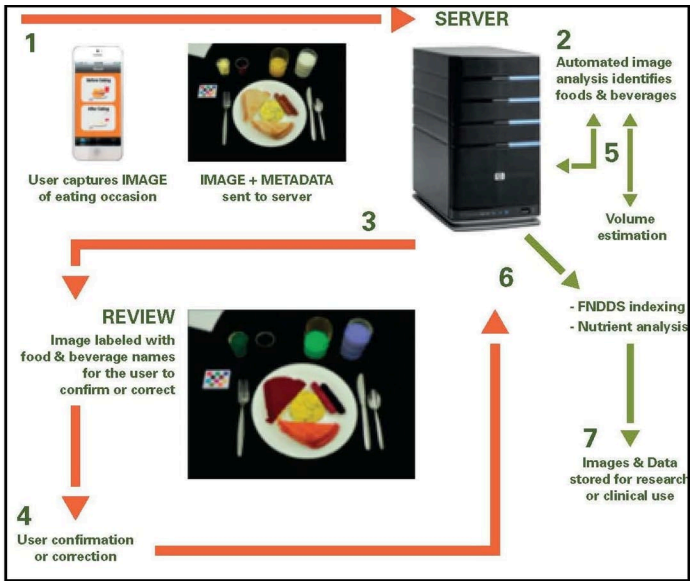


Figure 3.4 Diagram of the Technology Assisted Dietary Assessment (TADA) system. This shows (1) a user capturing an image of an eating occasion with the mobile food record; the image is sent to a server. (2) The image is analysed to identify the foods and drinks. (3) The labelled image is returned to the user for the review process; (4) The user confirms the automatic labels or corrects the labels. (5) The image is returned to the server for final identification and volume estimation. (6) Identified foods and amounts are matched for nutrient analysis to the Food and Nutrient Database for Dietary Studies (FNDDS). (7) Images and data are stored in a server for use by researchers or clinicians. Source: Dietary Assessment: A resource guide FAO, 2018.

Figure 3.5 compares the mean total energy expenditure (TEE based on doubly labeled water (DLW), with the reported energy intake (rEI) using images from a mobile food record (mFR) recorded over 7.5d by 45 community dwelling men and women, 21–65y (Boushey et al., 2017). In this study, mean energy intake for TEE and rEI were 2932 vs. 2353kcal/day with a significant mean difference of 580kcal/d, $p < 0.0001$). A moderate statistically significant correlation was found between rEI and TEE (Spearman correlation coefficients 0.58; $p < 0.0001$), with reporting accuracy that was consistent across all energy intake levels. The mFR was well received by the respondent and usability was rated as easy by Boushey et al.(2017, For details of the development of the mobile food record (mFR), see Six et al. (2010).

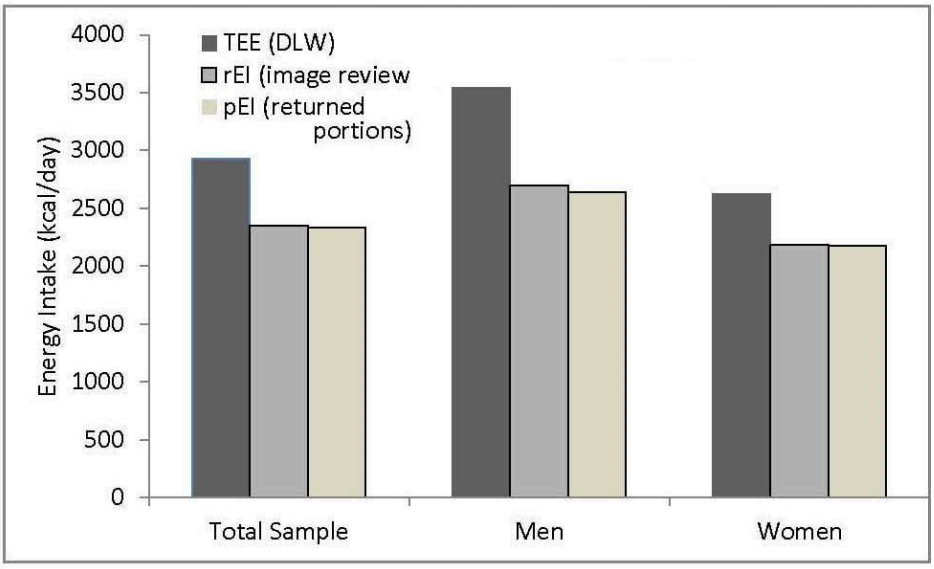


Figure 3.5. A comparison of mean total energy expenditure (TEE) based on doubly labeled water (DLW), reported energy intake (rEI) using images from a mobile food record, and presumed energy intake (pEI) based on returned preweighed servings of food over 7.5 days by total sample and by sex (men = 15 and women = 30). Modified from Boushey et al. (2017).

Unfortunately, there are few validation studies of image-based methods among the elderly, children, and adolescents, and the feasibility of their use in large studies (n > 100) has not yet been demonstrated. Burrows et al. (2019) provide a review of validation studies based on image-based methods using doubly labeled water.

3.2.5 Wearable devices

Under development are wearable devices worn around the neck, wrist, ear, or attached to clothing and designed to objectively measure diet without relying on user-reported food intake (Fontana et al., 2015). Some (e.g., WearSens, worn around the neck) lack the ability to detect specific types of foods, and thus quantify nutrient intakes, while others (e.g., GoBe, worn as a wristband) only measure calories, fat, protein, and carbohydrate. Several chewing and bite-related wearable sensors are also being developed which aim to differentiate between food intake and non-food intake, with some aiming to estimate energy intake (e.g., BitBite) (Fontana et al., 2015). Discomfort has been reported by some users of these wearable devices which suggests that their format needs to be modified (Boushey et al., 2017). Privacy issues also need to be addressed when using wearable devices (Gemming et al., 2015a).

This page titled 3.2: Technical improvements in food consumption measurements is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

3.3: Selecting an appropriate method

The method of choice for assessing food or nutrient intakes depends primarily on the objectives of the study. No method is devoid of random or systematic errors (Chapter 5), or prevents alterations in the food habits of the respondents. However, statistical techniques are now available which are designed to mitigate the impact of measurement errors on study results. See Freedman et al. (2011), Kirkpatrick et al. (2018), and Chapter 5. The US National Cancer Institute has developed a Dietary Assessment Primer (2020), a web resource to aid researchers choose the best available dietary assessment approach to achieve their research objective. Readers are advised to consult the primer before selecting a self-report dietary instrument.

Box 3.2 provides guidance on the most appropriate methods for assessing food or nutrient intakes in relation to four possible levels of objectives.

Box 3.2: Selection of methodology to measure nutrient intakes to meet four possible levels of objectives.

Level One: Mean nutrient intake of a group

- **Preferred approach** A single 24h recall, or single weighed or estimated food record, with large number of subjects and adequate representation of all days of the week

Level Two: Proportion of population "at risk"

- **Preferred approach** Replicate observations on each individual or a subsample using 24h recalls or weighed or estimated one day food records with an adequate representation of all days of the week

Level Three: Usual intakes of nutrients in individuals for ranking within a group

- **Preferred approach** Multiple replicates of 24h recalls or food records or a semiquantitative food frequency questionnaire

Level Four: Usual intakes of foods or nutrients in individuals for counseling or for correlation or regression analysis

- **Preferred approach** Even larger number of recalls or records for each individual. Alternatively, a semiquantitative food frequency questionnaire or a dietary history can be used.

Note that the number and selection of replicate 24h recalls, or weighed or estimated one day food records required to obtain level two, level three, or level four data, depends on the day-to-day variation within one individual (i.e., within-person variation) of the nutrient of interest (Chapter 6). This variation depends on the nutrient, the study population, and the seasonal variations of intake. Generally, for nutrients found in high concentrations in only a few foods, such as vitamins A and D and cholesterol, the number of replicates needed is greater than for those found in a wide range of foods (e.g., protein). Nonconsecutive days should be selected for the replicates when possible, to enhance the statistical power of the information: day-to-day correlations between intakes often occur when food intake data are collected over consecutive days. The length of time needed between the observation days also depends on the nutrient (IOM, 2000; a 3–10d interval after the previous recall is recommended (Tooze, 2020).

Additional factors that should be considered when choosing a method for assessing the food consumption of individuals are the characteristics of the individuals within the study population, the respondent burden of the method, and the available resources. For instance, certain methods are unsuitable for elderly subjects with poor memories, for busy mothers with young children, or for illiterate individuals. Other methods require highly trained personnel and specialized laboratory and computing facilities, which may not be available. Generally, the more accurate methods are associated with higher costs, greater respondent burden, and lower response rates. Unfortunately, compromises often have to be made between the collection of precise data on usual nutrient intakes of individuals and a high response rate.

3.3.1 Determining the mean nutrient intake of a group: level one

Level one is the easiest objective to achieve and can be met by measuring the food intake of each subject in the group using a single 24h recall or a one day food record, provided the individuals are representative of the study population and all the days of the week are proportionately represented in the final sample. Data on mean usual nutrient intakes of a group can be used for international comparisons across countries of the relationship of nutrient intakes to health and disease. However, this method based on a single intake day per subject should never be used for reporting the distribution of intakes (i.e., as percentiles of intake) (Tooze, 2020). To calculate n, the number of subjects required in the group, an estimate of the between-person variation for the nutrient of interest is needed. This is usually obtained from the literature but may be determined during a pilot study.

As an example, assume the expected mean iron intake obtained from the literature is 10mg/d, with an anticipated standard deviation (s) of 3mg/d. Also assume that we want to be 95% confident that the true mean lies between 9.2 to 10.8mg/d (i.e., the confidence interval has limits which are 0.8mg/d on either side of the mean). A 95% confidence interval is calculated approximately as the mean ±2 × e, where e is the standard error of the mean — a measure of the precision of the estimated mean. Hence , the required e = 0.8/2 = 0.4. We can use the following formula to calculate n, the desired group size.

$$n = s_b^2/e^2$$

where s_b^2 is the between-person variance of the nutrient of interest, and *e* is the desired standard error — a measure of precision required for the estimate of the mean intake of the nutrient of interest. Hence

$$n = 3^2/0.4^2 = 56.25$$

and so 56 subjects are required. Alternatively, if we wanted to be 99% confident that the true mean lies between 9.2 and 10.8mg/d, then more subjects must be studied. A 99% confidence interval is calculated approximately as the mean ±3 × e. Hence the required e = 0.8/3 = 0.27 Therefore

$$n = 3^2/0.27^2 = 123$$

and 123 subjects are required.

Clearly, the size of the group (n) necessary to characterize the group mean usual nutrient intake depends on the degree of precision required: more individuals must be studied to achieve higher precision. This calculation of the required group size should be repeated for each of the nutrients of interest, and the largest n (i.e., the worst case) should be used if possible.

If the study objective is to demonstrate a significant difference in the mean intakes of two groups, or a significant change in the mean intakes, based on unpaired or paired data, then alternative formulae must be applied; details are given in (Gibson and Ferguson, 2008).

3.3.2 Calculating the population percentage “at risk": level two

To determine the percentage of the population “at risk” of inadequate nutrient intakes, an estimate of the distribution of usual intakes of the individuals is required. This, in turn, requires that the food consumption of individuals be measured over more than one day. Hence, repeated 24h recalls, or replicate weighed or estimated one day food records are the methods of choice, again ensuring that all days of the week are proportionately represented in the final sample. Often, it is not feasible to carry out repeated observations on all the individuals, as in the case of a national dietary survey, and the recalls or records are repeated on a subsample of the individuals only.

To achieve a level two objective, at least two independent measurements of food intake should be obtained on at least a representative random subsample of individuals in the survey. The U.S. Food and Nutrition Board (IOM, 2000) recommends that the replicate measurements should be independent and made on non-consecutive days 3–10d after the first recall or record. However, if the data can be collected only on consecutive days, then three daily measurements should be used. The random subsample should consist of at least 50 individuals per demographic group. Where possible, the demographic groups to be sampled should be defined according to the sex and life-stage groups used for the chosen Nutrient Reference Values (Deitchler et al., 2020),

Note that it is more important to have a minimum number of replicate observations in the subsample than a minimum proportion of replicate observations. Once the required number of replicate observations have been obtained, methods can be applied to

correct for the measurement error associated with both day-to-day variability in intakes for a single person (termed within-person variation) and the variation in intakes between individuals (termed between-person variation) (see Chapter 6).

Four statistical methods are now available for estimating the distribution of usual intake. For more details, the reader is referred to (Tooze, 2020). Of the methods, the first was outlined by the National Research Council and later refined by Iowa State University (ISU) (Nusser et al., 1996) with software developed to use the ISU method. Note, this method does allow for a random sub-sample of repeated 24h recalls or food records, and has the capacity to adjust for season, day of the week, and/or sequence effects, but is not recommended for general use with episodically consumed foods, food groups, or nutrients (Tooze, 2020). The adjustment process provides estimates of the usual nutrient intakes for each specified age and gender-specific subgroup. An example comparing the adjusted distributions of usual zinc intakes (using the refined NRC approach of (Nusser et al., 1996) with the observed zinc intakes for New Zealand adult females aged 19–50y is shown in Figure 3.6.

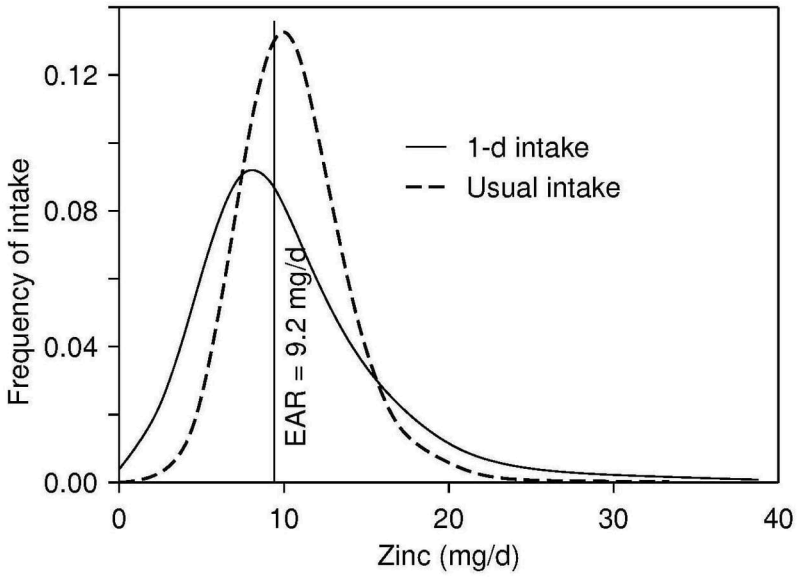


Figure 3.6 Estimates of usual intake distribution for zinc for New Zealand adults obtained from 24-hout recall data and adjusted with replicate intake data using the refined NRC method. The y-axis (frequency of intake) shows the likelihood of each level of intake in the population. EAR, Estimated Average Requirement. Modified from (Gibson et al., 2003). Nutrition today, 38(2), 63–70.

The adjustment process used yields a distribution with reduced variability (sometimes referred to as “shrinking”) because the within person variation has been removed, while preserving the shape of the original observed distribution (Gibson et al., 2003).

Later, the National Cancer Institute (NCI) (Tooze, 2020) developed a method which also adjusts for the same covariates (season, day of the week, and/or sequence), but also has the capability of incorporating covariates to identify estimates for sub-populations in the survey. In addition, the NCI method, unlike the ISU, can be used with episodically consumed foods, food groups, or nutrients.

The European Food Consumption Validation Project have developed the Multiple Source Method(Haubrock et al., 2011) (MSM) that can also be used for episodically consumed foods, food groups, or nutrients, although caution must be used when employing the program with models containing covariates. The Dutch National Food Consumption Survey 2007-2010 have developed a program called the Statistical Program to Assess Dietary Exposure (SPADE) , but this program currently requires 24h recalls on at least two-days for all respondents in the sample and is not recommended for use with episodically consumed foods, food groups, or nutrients.

All the methods described briefly above use varying approaches to yield an adjusted distribution of “usual" nutrient intakes that can then be used to predict the proportion of the population at risk of nutrient inadequacy using either the full probability approach, or the Estimated Average Requirement (EAR) cutpoint method; details are given in Chapter 8b. However, the choice of which “usual intake method” method to use depends on whether foods, food groups, or nutrients are episodically consumed, whether the probability of consumption and the consumption-day amount are correlated or not, and whether replicates of the 24h recalls or food records are available on all the population groups to be surveyed or only on a sub-sample (Souverein et al., 2011).

Table 3.7: "True" and observed prevalence estimates and the number of repeated measurements needed to reduce the observed prevalence to within 5% of the "true" prevalence. Data for women aged 45–54y from NHANES II(1976–1980). From: Sempos CT, Looker AC, Johnson CL, Woteki CE. (1991). The importance of within-person variability in estimating prevalence. In: Macdonald I (ed.), Monitoring Dietary Intakes, pp. 99–109. Springer-Verlag, Berlin..

Variable	Prevalence (%) "True"	Prevalence (%) Observed	No. of repeated measurements needed
Cholesterol > 300mg	15	37	39
Calcium > 800mg	12	21	9

Figure 3.6 shows that, in the example, adjusting the distribution significantly using the ISU method reduces the proportion of individuals considered to have intakes below the EAR for zinc. Within-person variation can also have a significant effect on estimates of the prevalence of abnormally high nutrient intakes. Table 3.7 shows data from NHANES II. The large differences between the observed prevalence and the calculated "true" prevalence represent the effect of removing the within-person variation by calculation. In this case, very large numbers of repeated measurements on each individual are required to reduce the observed prevalence of abnormally high intakes of cholesterol and calcium to within 5% of the true prevalence (Sempos et al., 1999). Comparable data are essential for national food policy development and food fortification planning. Food patterns associated with inadequate nutrient intakes can also be identified using this approach, enabling food assistance programs to be designed and improvements in nutrition education made.

More recently, a new method has been developed based on single-day dietary data which can be used to estimate population distributions of usual intake of nearly-daily consumed foods and nutrients, provided a suitable external within-person to between-person variance is available (Luo et al., 2019). Nevertheless, researchers are urged to collect replicate data where possible.

3.3.3 Ranking individuals by food or nutrient intake: level three

When the study objective is at level three and involves ranking individuals within a group, often for the purpose of linking dietary intakes with risk of chronic disease, the preferred approach is to obtain multiple observations on each individual. The number of days required to achieve the level three objective can be calculated from the ratio of the within- to the between-person variation in nutrient intakes (often termed the "variance ratio"); for more details see Chapter 6 . Sometimes an estimate of the variance ratio can be obtained from the literature, again preferably from an earlier study on a comparable group. Alternatively, a pilot study may be necessary to obtain this information.

Several authors have developed equations for calculating the number of replicate days required to meet level three objectives (Black et al., 1983; Basiotis et al., 1987; Nelson et al., 1989). Black et al. (1983) suggest using the following formula for the number of days (n) of diet records needed:

$$n = \left(r^2/\left(1-r^2\right)\right) \times \left(s_w^2/s_b^2\right)$$

In this equation, *r* is the unobservable correlation between the observed and true mean intakes of individuals over the period of observation, and s_w^2 and s_b^2 are the observed within- and between-person variances, respectively. This equation should be used in association with Table 3.8 which shows the proportion of individuals correctly and incorrectly classified in the extreme fractions for different values of the corrlation coefficient between the observed and true intakes (*r*). The value of *r* chosen will depend on the degree of misclassification that the investigator is prepared to accept (Table 3.8).

Table 3.8: Proportion of individuals correctly and incorrectly classified in the extreme fractions for different values of the correlation coefficient (r) . (a), correctly classified in the extreme thirds, fourths, or fifths of the distribution of intakes; (b), misclassified into the opposite extreme fraction. From Nelson et al. (1989).

		Correctly and incorrectly classified into extreme fraction		
r		Thirds	Fourths	Fifths

			Correctly and incorrectly classified into extreme fraction	
0.75	a b	0.69 0.049	0.63 0.013	0.59 0.004
0.80	a b	0.72 0.033	0.68 0.006	0.65 0.002
0.85	a b	0.76 0.018	0.72 0.002	0.69 <0.001
0.90	a b	0.80 0.006	0.77 <0.001	0.75 <0.001
0.95	a b	0.86 <0.001	0.84 <0.001	0.83 <0.001

As an example, assume that the investigator requires that when the individuals are divided into terciles, fewer than 5% (< 0.05) of the individuals are grossly misclassified into the opposite tercile. This will require an r value of 0.75 (Table 3.8). Assuming $s_w^2/s_b^2 = 1.7$, then the number of days (n)

$$n = \left(r^2/\left(1-r^2\right)\right) \times 1.7$$
$$n = 0.75^2/\left(1-0.75^2\right) \times 1.7$$
$$n = 3 \text{ days}$$

The number of days needed to generate a given r increases as the chosen r increases. If the size of the within-person variation (s_w^2) in nutrient intake is small compared with the size of the between-person (s_b^2) variation , then fewer replicate days are needed to meet level three objectives.

The U.S. subcommittee on criteria for dietary evaluation (NRC, 1986) recommended using independent days for replicating the measurements of one day nutrient intakes to reduce any effect of autocorrelation between intakes on adjacent days.

An alternative approach to achieving level three objectives is to use a semi-quantitative food frequency questionnaire. This approach is often used in epidemiological investigations to study associations between intakes and risk of disease and does not require a measurement of absolute nutrient intakes. Although this approach is much simpler, involving only a single interview with each subject, it is difficult to quantify the errors involved and to separate the effects of within- and between-person variance.

3.3.3 Determining usual intakes of nutrients of individuals: level four

Reliable estimates of usual food or nutrient intakes of individuals that can be used with confidence to meet a level four objective, involving correlation or regression analysis with individual biochemical measures, are the most difficult to obtain. Large numbers of measurement days for each individual are required using 24h recalls or estimated or weighed food records.

An estimate of the within-person variation for each nutrient of interest should be obtained from the literature, preferably from an earlier study on a comparable group or a pilot study, as noted earlier. This estimate may be expressed as the variance, s_w^2; standard deviation, s_w; or as the coefficient of variation (CV_w) expressed as a percentage:

$$CV_w = s_w/(\text{ mean intake }) \times 100\%$$

This estimate can be used in the following equation to determine the number of days required per individual to estimate an individual's nutrient intake to within 20% of their true mean 95% of the time (Beaton et al., 1979).

$$n = \left(Z_\alpha CV_w/D_0\right)^2$$

where n = the number of days needed per individual, Z_α = the normal deviate for the percentage of times the measured value should be within a specified limit (i.e., 1.96 in the example below), CV_w = the within-person coefficient of variation (as a percentage), and D_0 = the specified limit (as a percentage of long-term true usual intake) (i.e., 20% in the example given below).

The following example calculates the number of days required to estimate a Malawian woman's zinc intake using 24h recalls to

within 20% of the true mean, 95% of the time. In this example, the CV_w (i.e., 34%) for zinc intakes on Malawian women via 24h recalls is taken from the literature (IZiNCG, 2004). Thus if Z_α = 1.96 and CV_w = 34%. then:

$$n = (1.96 \times 34\%/20\%)^2 = 11d$$

If a pilot study is undertaken in which replicate 24h recalls are conducted, then the actual CV_w for each nutrient of interest can be calculated. In this way, the estimate of the number of days required to measure the usual intake of each of the nutrients of interest in an individual, with a required degree of precision, can be defined. In general, considerably more days are required to obtain reliable estimates of intakes of individuals to meet the level four objective, compared with level three (i.e., relative ranking of subjects into groups) (Palaniappan et al., 2003).

Sometimes, dietary histories or semi-quantitative food frequency questionnaires are used to obtain this level four data on usual nutrient intakes for correlation with biomarkers (Jacques et al., 1993). Some investigators emphasize, however, that the accuracy of a semi-quantitative food frequency questionnaire is only equivalent to two to three repeat 24h recalls (Sempos et al., 1999).

In some experimentally controlled studies such as balance studies, information on the actual nutrient intakes of an individual over a finite time period are required. For such data, weighed food records (Section3. 1.4), completed for the duration of the study period, are the recommended method. Nutrient intakes can then be calculated using food composition data. Alternatively, duplicate meals can be collected throughout the period for later analysis; details are given in Chapter 4.

3.3.4 Combining dietary instruments

Recognition of the limitation of even two 24h recalls to adequately measure, at the individual level, usual intake of foods or nutrients that are not consumed daily (i.e., episodically consumed foods) has led to the use of more than one dietary instrument in a given study. In this way, the advantages of each method can be exploited and their weaknesses minimized. Consequently, in large scale epidemiological studies, a common approach is the administration of a food frequency questionnaire to all respondents, and 24h recalls or food records in a subsample. The food frequency questionnaire provides information on foods that are consumed less frequently (often, “in the past year") not given by the 24h recalls while the 24h recalls yield richer detail by attempting to correctly quantify portion size for each eating occasion (Subar et al., 2006). Alternatively, a food frequency questionnaire could be used as a supplement to one or more quantitative 24h recalls administered to the full sample, providing richer detail and less systematic bias. This approach appears to have the most added value when the research question of interest is to estimate diet-health relationships, especially when the estimation of usual dietary intake at the individual level involves a food, food group, or nutrient that is consumed episodically (Tooze, 2020) Statistical methods have been developed to combine the dietary data from repeated 24h recalls and food frequency questionnaires. Conceptually, these statistical methods presume that the usual food intake of an individual equals the probability of consuming a food on a given day (queried as frequency of usual intake over a specified time period), multiplied by the average amount of intake of that food on a typical consumption day. Repeated 24h recalls from the same individual yield information on consumption probability and amount, while the food frequency questionnaire provides information on intake frequency of rarely consumed foods and usual intakes of nutrients when portion size information is available (Conrad and Nothlings, 2017). The reported food frequencies and nutrient intakes (if available) can be used as covariates in both steps of the NCI method (Tooze, 2020) and with MSM (Haubrock et al., 2011) to enhance the estimation of usual intakes from the 24h recall data, especially from foods that are not consumed every day. Note, however, that in low income countries where context-specific food frequency questionnaires are often not available or inappropriate in the study setting, it may be preferable to increase the number of repeat 24h recalls per respondent rather than embarking on the development, validation, and collection of dietary data using a food frequency questionnaire in addition to using the 24h recall method. For more details, the reader is advised to consult (Tooze, 2020). at: Intake.org

3.3.5 Surveys of individual food consumption at the global level

FAO/WHO Global Individual Food consumption data Tool (FAO/WHO GIFT) is a global database of surveys of food consumption at the individual-level which are collected at the national and subnational level throughout the world. Only data collected from 1980 by quantitative methods such as 24h recalls, food records (weighed or estimated), 12hr recalls, and direct food weighing are considered. Portion sizes of all foods and beverages consumed by each survey participant must be assessed, including water if possible. For low- or middle-income countries, surveys comprising at least 100 subjects (with no evidence of strong selection bias) are included, whereas for high-income countries, surveys that are nationally representative are an additional criterion. The dietary data generated from FAO/WHO GIFT can be used to inform agricultural, nutrition, food safety and environmental policies

and programs. A summary of 218 dietary surveys performed in low-and middle-income countries from 1980 to 2019 and included in the FAO/WHO GIFT inventory is available in de Quadros et al. (2022).

FAO/WHO GIFT apply several criteria to validate the dietary databases included. For example, for all foods and drinks consumed by each survey participant on each survey day, a complete description, the amount consumed, and energy and nutrient values for each item must be included in the datasets. Additional compulsory variables are the age and sex of each subject, geographical location (country, and region(s) if available), type of area (e.g., rural, urban), and the number of survey days recorded per subject or for a subset of subjects. Inclusion of variables based on anthropometric data such as weight and height and the physiological status of both women (e.g., pregnant, lactation) and infants (i.e., breastfeeding status) are also recommended.

Recipes should be disaggregated whenever possible to provide data on the quantity of each separate ingredient, cooking method used, total amount prepared, number of people served from the recipe, and the quantity consumed by the survey participant (or served and leftover). Only with this information can each ingredient be attributed to their appropriate food group for the calculation of the FAO/WHO GIFT indicators and summary statistics.

The food composition values used are provided by the data providers. However, all individual quantitative dietary datasets shared through FAO/WHO GIFT are coded with the FoodEx2 system, a comprehensive and flexible food classification and description system. Eligible dietary databases are mapped manually by the investigator with FoodEx2 codes for food groups (n=24) and food subgroups based on the descriptions provided by FAO/WHO GIFT. FoodEx2 was first developed by the European Food Safety Authority (EFSA) and was later scaled up to the global level in collaboration with FAO and the World Health Organization (WHO). The use of this common food classification and description system among dietary surveys from different countries contributes to the global harmonization of dietary data. Prior to the analysis and formatting by FAO/WHO GIFT, all eligible datasets are screened for potential errors, missing values and outliers. For further details see FAO/WHO GIFT (Methods)

Provided the eligible dietary datasets have been coded using FoodEx2, the FAO/WHO GIFT platform can compute ready-to-use indicators and summary statistics based on individual food consumption data in the areas of food consumption, food safety and nutrition. For example, of the indicators, one reflecting dietary diversity at the population level (a key component of diet quality) entitled the Minimum Dietary Diversity for Women(MDD-W) is computed. This is a food group-based indicator that estimates the proportion of non-pregnant women of reproductive age (15-49y) who consumed at least five out of ten defined food groups over the previous 24h.

The summary statistics generated by FAO/WHO GIFT comprise the estimated usual intakes of selected nutrients for pre-defined population groups by sex and age (except children less than aged 12 months) provided the datasets contain multiple non-consecutive days of 24h recalls / food records for at least a subset of 50 individuals. The statistical program chosen to adjust the dietary data for day-to-day variation is the Statistical Program to Assess Dietary Exposure (SPADE) by Dekkers et al. (2014). In addition, information on Nutrient Reference Values (NRVs) set by FAO/WHO, the European Food Safety Authority (EFSA) and the Institute of Medicine (IOM) are provided on the FAO/WHO GIFT platform to facilitate comparison of the estimated usual intakes of selected nutrients with NRVs. For more details of the statistical program, see (SPADE) and the instruction manual (SPADE - 4100). For NRVs and their applications, see Chapters 8a and 8b, respectively.

The Global Dietary Database (GDD) aims to identify, compile, and standardize individual-level data on dietary factors related to maternal-child health and chronic diseases. Currently the GDD comprises surveys across 188 countries conducted between 1980 and 2018 and provides empirical evidence on dietary intakes both across and within countries worldwide. Priority is given to nationally or sub-nationally representative dietary surveys based on 24-h recalls, food frequency questionnaires or short standardized questionnaires (e.g., Demographic Health Surveys (DHS). Household-level surveys are included if individual-level surveys are not available in a country and converted to individual-level intakes within each household using Adult Male Equivalents (AME), also known as Adult Consumption Equivalents. See: Weisell & Dop (2012). These authors account for the household composition and differing energy intakes by age and sex of household members. See Coates et al. (2017) for discussion of the validity of the application of the AME method for household-level data.

The FoodEx2 categorization system is used to standardize the description and classification of foods into food groups. Standardization also includes categorising nutrients and their units; quality assessment; aggregation by demographic strata and energy adjustment. For more details on data extraction and standardization, see: (GDD)

Mean intakes of 54 dietary factors by country, year, age, sex, education, urbanicity, and pregnancy / lactation status within nations can now be estimated using the GDD prediction model in 188 countries / territories in Asia, Asia-Pacific high-income countries, Oceania, Former Soviet Union, Latin America and Caribbean, Middle East and North Africa, South Asia, Sub-Saharan Africa, and the Western high-income countries in Australasia, Europe and North America (Miller et al., 2021). The dietary factors selected

and defined based on evidence for relationships with maternal-child health or chronic diseases include 14 foods, 7 beverages, 12 macronutrients, and 18 micronutrients. See (GDD) for more details on data extraction and the standardization used for estimating dietary intakes. Several indicators of global dietary patterns (e.g., Alternative Healthy Eating Index (AHEI); Dietary Approaches to Stop Hypertension (DASH), and the Mediterranean Diet Score (MED)) among children and adults have also been compiled from the GDD sets and compared globally, regionally, and nationally (Miller et al. ,2022).

Institute for Health Metrics and Evaluation (IHME) initiative database aims to provide rigorous and comparable measurement of the world's most important health problems and evaluates the strategies used to prevent them. IHME uses FAO Food Balance Sheet estimates, national product sales, household surveys, and data based on 24h recalls (considered the gold standard). Datasets created by IHME are stored in the IHME data catalogue known as the Global Health Data Exchange and can be freely downloaded from the IHME website.

IHME provides freely available modeled data by country, age, sex and year, based on primary data collected in 204 countries and 87 indicators, including 15 dietary indicators (9 foods and 6 nutrients). The dietary indicators are included in GBD 2017, a worldwide observational epidemiological study that tracks the progress within and between countries of the changing health challenges (Lim et al., 2013).

This page titled 3.3: Selecting an appropriate method is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

4: MEASUREMENT ERROR IN DIETARY ASSESSMENT (CHAPTER 5)

Abstract

Data on food consumption are typically collected using self-report methods. The resulting data are affected by measurement errors that can have serious consequences for study findings and interpretation. Measurement error refers to the difference between true and observed intake and may be random or systematic. Errors arise due to the interaction of the participant with the assessment method and can also be generateded by interviewers and coders, as well as by limitations in food composition databases. Accordingly, the type and extent of the errors vary with the method used and how it is implemented, the target population of interest, and the nutrients and foods investigated.

Both unaddressed random and systematic measurement error can introduce substantial bias into results. Pertinent to surveillance and monitoring, measurement error can lead to erroneous inferences about the proportions of a population with inadequate or excessive intakes relative to nutrient requirement estimates and food group recommendations. In epidemiologic research, measurement error distorts observed associations between diet and disease, as well as reducing statistical power to detect such associations. In intervention research, measurement error can mask the effects of the intervention, particularly if the error is differential between intervention and control groups. Strategies to minimize and/or mitigate error are therefore fundamental to research making use of dietary intake data and should be considered early in study design through to reporting results and implications.

4.1: Measurement error in dietary intake data

4.2: Sources of measurement error

4.3: Implications of measurement error for estimated dietary intakes

4.4: Minimizing measurement error through data collection procedures

4.5: Summary

This page titled 4: Measurement Error in Dietary assessment (Chapter 5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Measurement error in dietary intake data

Sources of both random and systematic error in commonly used methods (Chapter 3) for measuring dietary intake are discussed in this chapter. Chapters 6 and 7 discuss the related concepts of reproducibility and validity. Error associated with the compilation of nutrient composition data and the nutrient analysis of food items are discussed in Chapter 4. Error introduced by inappropriate study design and/or sampling of the target population are covered in detail in survey methods and epidemiology texts, and considered briefly in Chapter 1.

In the context of dietary assessment, measurement error refers to the difference between true and observed intake and may be random or systematic. Data affected by random error only are not biased but they are imprecise. Accordingly, random measurement error generally affects the reproducibility (Chapter 6) of a given method; that is, the extent to which the method yields similar results when used repeatedly in the same situation. A key source of random error is within-person variation in intake over time. Therefore, random error is larger in data collected using methods that capture intake over the short term, or quantitative daily consumption methods (Chapter 3), including 24h dietary recalls and food records, compared to those like food frequency questionnaires that aim to capture intake over a longer period, such as a year (Kipnis et al., 2003; Freedman et al., 2014, 2015; Kirkpatrick et al., 2022b).

The random error in short-term data is largely driven by within-person day-to-day variation in intake. That is, what individuals eat and drink changes from day to day, to different extents depending on the nutrient and food group of interest, as well as the context and food supply. Within-person variation in intake is not itself a form of bias since intake truly varies from day to day. However, random error must be considered in data analysis and interpretation because it results in an observed distribution of intake for a population based on data for a day, or even a few days, that is too wide relative to the distribution of usual intake (Figure 5.1).

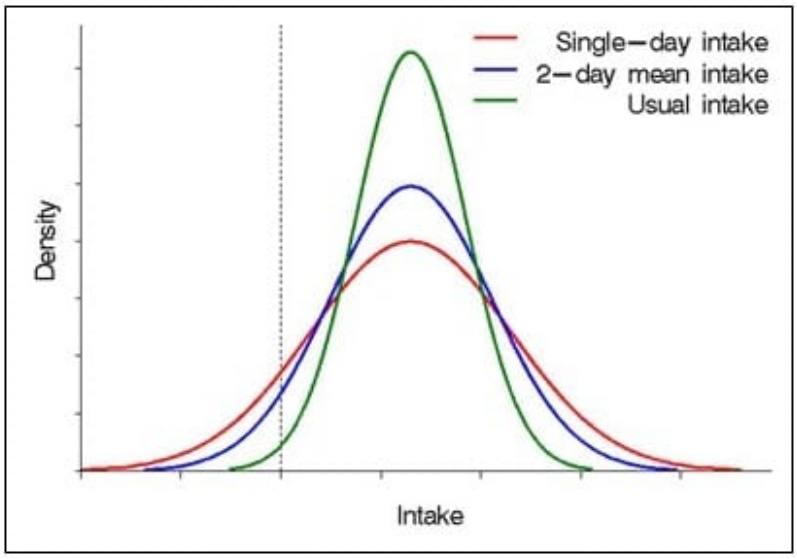


Figure 5.1.

Accordingly, when estimates such as the proportions of a target population with intakes above or below some threshold, such as food group recommendations, are based on an observed distribution affected by within-person variation, they will be biased. To estimate distributions of usual intake for a population, repeat measures on at least a subsample and statistical modeling techniques can be used to adjust for, or remove, the within-person variation (see Chapter 3). The use of short-term data to make inferences about usual intake is discussed in detail by Kirkpatrick et al. (2022a).

Systematic error affects data from all self-reported dietary assessment methods. However, systematic error, also known as bias, is typically larger in data collected using methods like food frequency questionnaires that aim to capture intake over a longer period, such as a month or year (Kipnis et al., 2003; Freedman et al., 2014, 2015; Kirkpatrick et al., 2022b). When the

observed distribution of intake, including the mean, is affected by systematic error, it is shifted left or right compared to the true distribution of intake. Therefore, estimates of mean intake and other parameters of the distribution, including proportions above or below thresholds such as food group recommendations, are biased. Systematic error generally affects the validity of a method. In epidemiologic research, measurement error distorts observed associations between diet and disease, as well as reducing statistical power to detect such associations (Figure 5.2).

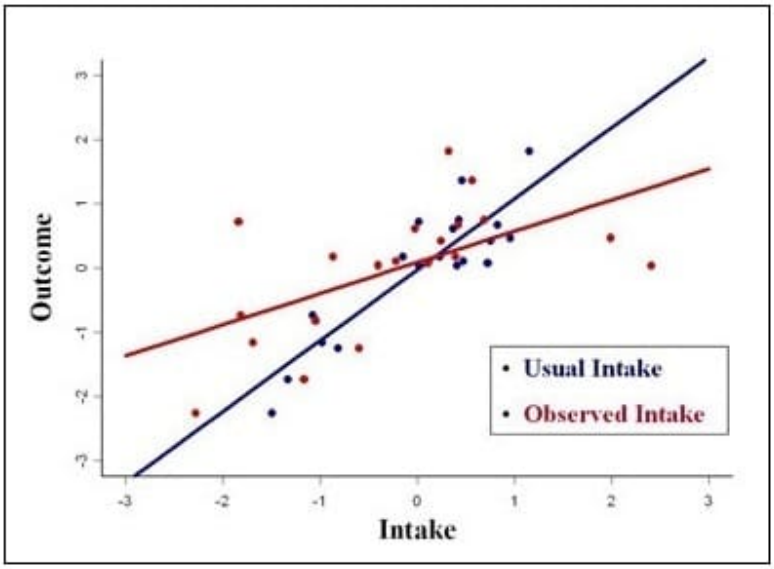


Figure 5.2.

Systematic measurement error is much more difficult to mitigate than sources of random error and unlike the latter, cannot be addressed using repeat measures. Rather, it is possible to correct for systematic error using statistical modeling only in situations in which data collected using an unbiased reference measure, such as a recovery biomarker, are available for at least a subsample, for example, from a calibration substudy. In cases in which data are available for at least a subsample from a less biased but error prone rather than an unbiased method, it may be possible to somewhat mitigate the impacts of systematic measurement error. The implications of systematic error for the implications of findings must be considered in all research making use of self-report dietary intake data (Freedman et al., 2011; National Cancer Institute, 2015; Subar et al., 2015).

This page titled 4.1: Measurement error in dietary intake data is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

4.2: Sources of measurement error

Reporting dietary intake is a complex cognitive process(Smith, 1993; Baranowski & Domel, 1994; Nelson et al., 1994). In addition to within-person variation and other random errors, such as random misestimation of amounts consumed caused by rounding, major sources of errors driven by the interaction between respondents and the assessment method include recall, social desirability, and reactivity biases. In studies in which dietary assessment methods are interviewer-administered, error may be introduced if different interviewers probe for information to varying degrees, intentionally omit certain questions, or record responses incorrectly. In studies using web-based or other technology-enabled methods, error may come about due to a mismatch of the interface with the cognitive abilities, such as literacy and numeracy abilities, of the target population. Another potential source of misalignment is a lack of inclusion of foods and beverages commonly consumed by the target population, such as in the food lists used in web-based 24h recalls and in food frequency questionnaires. Errors can also be introduced during coding, including when foods and beverages are assigned codes, including in systems using automated coding, and when portion sizes are converted from household measures, such as measuring cups, into grams.

5.2.1 Recall bias

Complete and accurate reporting of dietary intake using both short-term/daily consumption and long-term consumption methods can be impacted by recall errors. That is, individuals may not accurately remember what was consumed, along with the corresponding details, over either the short term (specific memory) or the long term (generic memory) (National Cancer Institute, 2015).

The 24h recall, which prompts participants to report retrospectively what was consumed on the prior day from midnight to midnight or during the previous 24h period (Section 3.1.1), relies largely on specific memory (National Cancer Institute, 2015). Imperfect memory, or recall bias, may result in the omission of eating occasions, foods, beverages, and supplements. Imperfect memory may also lead the participant to report foods that were not consumed during the recalled day (error of commission or intrusions) (Guinn et al., 2008; Baxter et al., 2009b, 2015). Both omissions and intrusions have been observed in studies in which 24h recalls were compared with observed intake based on unobtrusive recording on the same day (Krantzler et al., 1982; Karvetti & Knuts, 1985; Brown et al., 1990; Kirkpatrick et al., 2014b, 2019; Baxter et al., 2015). Recall bias can also result in the omission of or inaccuracies in details, such as additions to foods and beverages (e.g., honey in tea) and amounts consumed. In general, food items that contribute significantly to the main part of a meal are better remembered than additions such as condiments and salad dressings. In early research assessing adults' ability to report amounts of foods consumed, Guthrie (1984) reported that for one in six respondents, salad dressings were forgotten. A more recent study to assess the validity of the web-based Automated Self-Administered 24-Hour Dietary Assessment Tool (ASA24) similarly found that omissions were mainly additions to or ingredients in multicomponent foods, such as vegetables and condiments in salads and sandwiches(Kirkpatrick et al., 2014b). In that study, fruits and vegetables were omitted to a greater extent than sweets, snacks, and desserts, potentially because fruits and vegetables were mainly offered as part of other dishes. Table 5.1 shows the most common items that were truly consumed based on observation but were not reported on recalls completed using ASA24 and administered by interviewers using AMPM.

Table 5.1 Counts of most common exclusions, by recall mode (ASA24 and AMPM), in relation to true (observed) intakes.

Items	ASA24	AMPM
Tomatoes	42	26
Mustard	17	17
Green and/or red pepper	16	19
Cucumber	15	14
Cheddar cheese	14	18
Lettuce	12	17
Mayonnaise	9	12

Chapter 7 further discusses the results of studies assessing the accuracy of 24h recall data. In food records, respondents may likewise omit some foods and beverages consumed, including from images taken when using mobile device-based records.

To facilitate more accurate recall of food consumption, automated multiple-pass 24h recalls are used in many national surveys and studies. Multiple-pass interviewing techniques (Section 3.1.1) include "probing" questions, standardized "prompts," and/or memory aids such as food models. The multiple-pass approach is intended to minimize the omission of possible forgotten foods and standardize the level of detail for describing common foods, as well as the methods used to elicit specific details for certain food items. In the EPIC study, a computer-assisted 24h diet recall method using the software program EPIC-SOFT was developed to standardize the cognitive memory aids used in the first stage (called the quick list) of the multiple-pass recall (Slimani et al., 1999; Crispim et al., 2014). EPIC-SOFT, now known as GloboDiet, was further developed to support its use in pan-European and other international surveys (Slimani et al., 2011; Crispim et al., 2014), and hasbeen adapted for use in other contexts (Aglago et al., 2017; Bel-Serrat et al., 2017; Steluti et al., 2020). Automated multiple-pass methods may be implemented on smartphones and tablets to improve feasibility, including in low-income settings with limited electricity and connectivity(Caswell et al., 2015; Harris-Fry et al., 2018; Htet et al., 2019; Rogers et al., 2022).

The interviewer-administered Automated Multiple-Pass Method (AMPM) was developed by the USDA (Blanton et al., 2006; Moshfegh et al., 2008; Steinfeldt et al., 2013a; Rhodes et al., 2013). (Section 3.1.1) and is used in the US NHANES, with adapted versions used in the Canadian Community Health Survey (Health Canada, 2006, 2017) and the Australian Health Survey (Australian Bureau of Statistics, 2015). The AMPM was adapted for self-administration within the Automated Self-administered 24h Dietary Assessment Tool (ASA24) (Subar et al., 2012), with the similar goal of facilitating more complete reporting and reducing omissions, for example, through a forgotten foods list and repeated prompts to report everything consumed on the prior day or during the previous 24h period. To facilitate self-administration, ASA24 starts by asking respondents to report eating occasions (i.e., meals and snacks) and then prompts for the foods and beverages consumed at each one. Other countries have begun to integrate web-based recall systems into their national surveys; for instance, starting in 2019, dietary intake data for the United Kingdom's National Diet and Nutrition Survey have been collected using Intake24 (Simpson et al., 2017; Office for Health Improvement & Disparities 2023). Minimizing the period between food intake and its recall, or the retention interval for recall methods (Baxter et al., 2004), may reduce respondent memory lapses (Smith et al., 1991). Research related to the retention interval has focused on children and has demonstrated better accuracy with a shorter retention interval (Baxter et al., 2004, 2009a, 2009b). Automated systems such as ASA24 and the Nutrition Data System for Research (NDSR), a computerized multiple-pass protocol used to collect dietary intake data in the United States, (Nutrition Coordinating Center; Schakel et al., 1988), provide options for the researcher to administer recalls for the prior 24h rather than the prior day from midnight to midnight (Nutrition Coordinating Center; National Cancer Institute, 2024a).

Probing questions and standard "prompts" also reduce memory lapses. For example, early research suggested that probing increased reported dietary intakes of hospital patients assessed by 24h recalls by 25% compared to intakes obtained without probing (Campbell & Dodds, 1967). In a U.S. study of fourth-grade school children, the accuracy in reporting foods eaten for school lunch 90min earlier increased to 100% after responding to prompts for additional items (Domel, 1997). Baxter et al., (2015) have examined the impact of different types of prompts, including open prompts that ask respondents to report intake freely (used by AMPM), meal-name prompts (used by ASA24), forward prompts (respondents are instructed to report consumption from the beginning to the end of the reporting period; used by NDSR), and reverse prompts (respondents are instructed to report the most recent consumption first) on the accuracy of reported intake among fourth-grade children. The authors found that the prompting method interacted with the retention interval; when the retention interval was short (with the analysis focused on a recent period within the prior 24h), there was no apparent difference in accuracy in relation to the type of prompting. When the retention interval was longer (with the analysis focused on the morning of the prior day), reverse prompting was associated with higher accuracy (Baxter et al., 2015). Depending on the target population, combinations of retention interval and prompting may be used to facilitate recall; however, manipulations to prompting may not be possible using automated systems and may impact comparability with other studies.

In addition to lessening the retention interval and improving prompting, shorter reporting periods have been considered to improve accuracy of recall. Lucassen et al. (2023) observed slightly more accurate estimates of protein and potassium intake based on smartphone-based 2h versus 24h recalls when compared with urinary biomarkers (Lucassen et al., 2023). Additionally, most participants in the study preferred the 2h recall compared to 24h recalls administered using either a web-based interface or via interviewers. It is possible that multiple 2h recalls spread across days and times of day could be used to estimate intake somewhat more accurately than repeated 24hrs while reducing participant burden.

Portion-size measurement aids (PSMAs) can also help facilitate accurate recall. Aids may be photographic or non-photographic (Section 3.2.2) and range from 1-dimensional to 3-dimensional (Amoutzopoulos et al., 2020). Aids available as a range of graduated portions may reduce portion size measurement error. In such cases, a series of graduated food models or photographs of a range of portion sizes can be used. Systems such as ASA24 and NDSR also integrate recipe features with the aim of capturing dietary intake more accurately (Nutrition Coordinating Center (NCC) n.d.; National Cancer Institute 2024a).

Data from 24h recalls can be used to examine contextual details, such as with whom meals were eaten, when, and where. These details may be impacted by imperfect recall. Most validation research focuses on the accuracy of quantitative estimates of intake (Chapter 7) versus these contextual details. Given increasing attention to meal patterning (Leech et al., 2015a), as well as ongoing interest in the influence of food environments on dietary intake (Kirkpatrick et al., 2014a), additional attention to the accuracy with which these details are reported may be warranted.

Intake reported using methods like the food frequency questionnaire that prompt participants to retrospectively report intake over a longer period, such as a year, are also reliant on memory. Probes are also useful for this method, which relies mainly on generic versus specific memory. Nonetheless, the cognitive demands associated with remembering what was consumed can be exacerbated by the requirement to average consumption of a given food over a long period. Frequency questionnaires may also require respondents to consider multiple similar foods captured by one line item on the questionnaire that may be consumed with varying frequencies. Estimating average consumption over time may be especially difficult for foods that vary in availability according to season. Some food frequency questionnaires therefore include questions related to seasonal food consumption (Section 3.1.4) to reduce recall bias.

Food records (Section 3.1.2) completed in real time-that is, at the time of consumption-may be subject to some recall bias if details, such as the level of fat or type of sweeteners in a food in a packed lunch, are not accurately remembered. If records are completed at the end of the record keeping period rather than in real time, they become like 24h recalls in terms of relying on retrospective reporting and can be affected more substantially by recall bias.

Recall bias is not consistent across individuals and may be more pronounced among children who have not yet fully developed their cognitive abilities, including memory, and individuals who have cognitive limitations. For this reason, it is important to carefully tailor the assessment method used with the target population; for instance, young children are unlikely to be able to accurately complete a 24h recall on their own and proxy- or proxy-assisted reporting may be needed.

5.2.2 Social desirability bias

Reported intake may be affected by social desirability biases, the tendency to give responses perceived as socially desirable, and social approval biases, the tendency to seek praise (Miller et al., 2008). For example, prompting during 24h recalls and food lists within food frequency questionnaires that focus on foods perceived as desirable or undesirable may elicit this source of bias. Social desirability bias may be intentional or a form of self-deception (Roth et al., 1986).

Worsley et al. (1984) suggested that reported intakes of certain foods such as fresh fruits and vegetables and sweet foods are particularly susceptible to social approval needs and hence are a potential source of systematic bias through over- or under-reporting. With increasing attention to limiting consumption of certain foods such as sources of free and added sugars (WHO, 2015) and highly processed foods (Ministry of Health of Brazil, 2015), these items may be prone to increasing social desirability bias over time, with implications for monitoring trends in intake and evaluating the impact of dietary guidance and other policy initiatives. Relatedly, social desirability bias may be higher among persons living in larger bodies given the stigmatization of overweight and obesity in many societies (Brewis et al., 2011; Puhl et al., 2018). If the study population is stratified by weight status based on body mass index, differences in reporting may mask differences in intake (Beaton et al., 1997; Lissner et al., 2007). Although prior work has suggested greater social desirability and social approval biases in women (Hebert et al., 1995, 2008), it is possible this is shifting given changes in social norms and increasing attention to body weight, thinness, and muscularity across individuals with different gender identities (Puhl et al., 2018; Himmelstein et al., 2018; Nagata et al., 2021).

Hebert et al. (2008) used a 10-item version of the Marlowe-Crowne Social Desirability Scale (Table 5.2) and found that the largest social desirability biases related to reporting fruit and vegetable intake using a short frequency-based screener were among male and female participants with low educational attainment. The researchers also found that the bias was not consistent over time, with women randomized to an intervention reporting intake with higher bias at follow-up versus baseline, whereas bias among men was higher at baseline. Miller et al. (2008) also found substantial social approval bias affecting data collected from women using a short frequency questionnaire and a limited 24h recall focused on fruit and vegetable intake. Participants were exposed to messaging about the benefits of and guidelines for consumption of fruits and vegetables.

Table 5.2 Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. Strahan & Gerbasi, 1972

	T or F	Statement
2.	(T)	I never hesitate to go out of my way to help someone in trouble.
4.	(T)	I have never intensely disliked anyone.

	T or F	Statement
20.	(T)	When I don't know something I don't at all mind admitting it.
21.	(T)	I am always courteous, even to people who are disagreeable.
24.	(T)	I would never think of letting someone else be punished for my wrong doings.
6.	(F)	I sometimes feel resentful when I don't get my way.
12.	(F)	There have been times when I felt like rebelling against people in authority even though I knew they were right.
14.	(F)	I can remember "playing sick" to get out of something.
28.	(F)	There have been times when I was quite jealous of the good fortune of others.
30.	(F)	I am sometimes irritated by people who ask favors of me.

Similarly, analyses of baseline data collected from women enrolled in the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and participating in a nutrition education intervention related to fruits and vegetables found that higher social desirability trait, measured using a short version of the Marlowe-Crowne Social Desirability Scale, was associated with higher reported frequency of consumption of fruits and vegetables but only among women who were not breastfeeding (Di Noia et al., 2016). These studies used short tools focused on foods that may be particularly affected by social desirability bias (i.e., knowledge about the benefits of and recommendations for fruit and vegetable consumption are widespread). Additional studies considering other types of foods and beverages would be beneficial for expanding knowledge in this area.

Data affected by social desirability bias can be particularly problematic for monitoring the effectiveness of interventions, especially those focused on changing individual behaviors, such as through counselling or educational programs (Buzzard & Sievert, 1994; National Cancer Institute, 2015). Some research has demonstrated changes in dietary reporting on 24h recalls and food frequency questionnaires due to exposure to an intervention (Kristal et al., 1998; Caan et al., 2004; Natarajan et al., 2010). When an intervention group is compared to a control group, this differential reporting-that is, reporting error that is different in the intervention group compared to the control group-may mask the effects of the intervention.

Collaborating with cognitive psychologists to expand understanding of the implications of social desirability and social approval biases, along with other psychosocial factors, may inform strategies to modify methods and/or analytic methods to reduce the impacts of these sources of measurement errors. The use of a social desirability scale in dietary surveys to identify and perhaps control for social desirability variables has been suggested (Worsley et al., 1984; Hebert, 2016), but is not yet a widespread practice. The incorporation of measures of social desirability in validation studies that collect unbiased reference data (Di Noia et al., 2016; Hebert, 2016) could expand our understanding of the magnitude of this source of bias and the populations and dietary components for which it is most problematic.

Given the potential impact of social desirability bias on capacity to evaluate interventions, data aside from or in addition to self-reported intake are recommended. Depending on the intervention, such data may come from direct observation, biomarkers, or food purchasing records, though to date, there has been relatively little attention to analytic methods to combine such sources of data to evaluate an intervention (Miller et al., 2008; Keogh et al., 2016; Kirkpatrick et al., 2018)

5.2.3 Reactivity bias

In contrast to methods that query intake retrospectively, those that involve real-time recording, such as food records, can elicit reactivity. Reactivity is defined as a change in behavior due to awareness that behavior is being or will be measured (Section 3.1.2). The respondent may modify their food consumption during the reporting period, potentially in a more socially desirable manner and/or to simplify the recording task (National Cancer Institute, 2015; Burrows et al., 2019).

To minimize reactivity, 24h recalls are generally intended to be unannounced such that the respondent does not know they will be asked to complete the recall on a given day. In some web-based recall systems, a scheduling option allows for researchers to invite respondents to complete their recalls on given days and to prevent recalls from being completed on other days. In situations in which the respondent knows that a recall is upcoming or can choose the day on which to complete a recall, reactivity bias may be a nontrivial contributor to the overall systematic error.

For food records, reactivity may be a desired outcome of self-monitoring as part of an intervention aimed at changing eating practices and/or body weight (Streit et al., 1991; Tinker et al., 2001). However, when interventions such as counselling are being evaluated for their ability to change intake, food records are not recommended for evaluating the intervention (National Cancer Institute, 2015), as discussed in Section 5.2.2, because of challenges in differentiating changes in reported intake due to the intervention from changes in reported intake due to reactivity.

In studies with repeated administrations of dietary assessment methods, it is possible that reactivity may decline over time as the novelty of monitoring wears off. However, it is challenging to distinguish changes in this source of bias from other biases and factors, such as practice effects, that may be dynamic across repeat administrations.

5.2.4 Interviewer biases

Errors in reported dietary intake can also come about due to interviewer biases (Wynder, 1994; Davis et al., 2010). These biases are most pertinent to 24h recalls, which continue to be interviewer administered in many studies and national surveys (De Keyzer et al., 2015), though this is changing, as noted in Section 5.2.1. Interviewer biases can be random across days and respondents, and/or systematic for a specific interviewer, or they may exist as an interaction between certain interviewers and certain respondents (Anderson 1986).

Interviewer biases may include errors caused by incorrect use of probing questions, incorrect recording of responses, intentional omissions, biases associated with the interview setting, and distractions (Fowler and Mangione 1990). The degree of rapport between the interviewer and the respondent can also play a role, for example, lending to the extent to which a respondent exhibits social desirability bias. Relatedly, mode of administration is relevant, with auditory and visual cues in face-to-face interviews potentially leading to larger interview bias compared to telephone interviews (Davis et al., 2010). Interviewer identities, including racial/ethnic and gender identity, can also play a role, as can the use of interviewers with limited characteristics (e.g., only women) (Davis et al., 2010). A large and diverse interviewing staff, along with thorough training and monitoring, are recommended (Davis et al., 2010).

A carefully designed and standardized interviewing protocol, preferably computer administered, may help minimize the effect of interviewer biases(Slimani et al., 2000; Vossenaar et al., 2020). When several interviewers are employed, the assignment of interviewers-respondents-days should be randomized, and the interviewers should be trained to anticipate and recognize potential sources of distortion and bias (Wakefield 1966). In situations in which interviewers are conducting dietary assessment using online interfaces, including those designed for self-administration by respondents, thorough and consistent training is recommended to ensure consistency with the protocol and across interviewers. In a study by Baxter et al. (2015), researchers who did not conduct interviews reviewed the interview audio recordings and transcripts to assess adherence to the protocol.

For dietary surveys involving diverse groups, it is advisable to use interviewers familiar with each language and culture. However, care must be taken to ensure that individual interviewers are not assigned only to specific subgroups of the target population, as interviewer and group effects can become confounded. Special attention must be given to the way the questions are asked, and how people think about and describe foods and amounts consumed in each ethnic or cultural group (Hankin & Wilkens, 1994). This will necessitate training in focused ethnographic methods (Buzzard & Sievert, 1994). The Intake Center for Dietary Assessment, in guidance for conducting large-scale 24h recalls in low- and middle-income countries, recommends a minimum of three weeks of training to practice dietary data collection, in addition to a formal survey pilot and final field testing (Deitchler et al., 2020). Surveys should be piloted in each language (Deitchler et al., 2020).

Comparing nutrient intakes calculated from multiple interviews carried out independently on the same participants during the same 24h eating period, using different trained interviewers, may reveal interviewer effects. Despite rigorous efforts to standardize a computer-assisted 24h recall method used in the European prospective investigation into cancer and nutrition (EPIC), an interviewer effect was observed in certain countries among the 90 interviewers involved in the collection of approximately 37,000 24h recalls (Slimani et al., 2000), although not across centers from the same country. In this study, the interviewer effect was based on mean energy intake per interviewer, considering the confounding effects of age, body mass index, energy requirements, weekday, season, special diet, and physical activity. Assessment of potential interviewer bias in studies making use of multiple interviewers can allow for statistical methods to be applied to correct for this source of dietary measurement error (Slimani et al., 2000), such as by using models that account for clustering of respondents by interviewers (Davis et al., 2010).

5.2.5 Coding errors

Processing food consumption data to enable estimation of intake of energy, nutrients, and other dietary components involves coding each item reported using a food composition database (Chapter 4) and depending on the nature of the data, the quantity (portion size) or frequency of intake (Guan et al., 2017). Coding may be an overlooked potential source of error.

Accurate coding can be facilitated by collecting detailed information about foods and beverages consumed, for example, using multiple-pass methods (Sections 3.1.1 and 5.2.1). Beaton et al. (1979) reported that coding error arising exclusively from inadequate descriptions of foods rather than weight error resulted in coefficients of variation ranging from 3% for protein and 8% for total fat to 17% for the ratio of polyunsaturated to saturated fatty acids. Coding is also challenging if the reported foods or beverages are not available in the database, especially for open-ended data collected using a method such as a diet history questionnaire (Guan et al., 2017), requiring judgement on the part of the coder. Discrepancies among coders are a source of error (Guan et al., 2017), and such discrepancies over time and across settings can confound potential changes or between-context differences in dietary intakes (Buzzard & Sievert, 1994; Slimani et al., 2000). Additionally, databases often provide codes for composite or generic foods that represent many potential variations, posing challenges for accurately coding. However, the availability of brand-specific food composition databases is increasing (Carter et al., 2016; Poti et al., 2017; Westenbrink et al., 2021).

Errors in coding can be reduced if "coding rules" are established to deal with incomplete or ambiguous descriptors of the foods (Anderson 1986). and databases with a comprehensive range of food items are used to enable coders to select the appropriate food item more readily. Maintaining databases that reflect the current food supply is an ongoing challenge (Faber et al., 2013). To promote adequate and reliable information on food composition, the International Network of Food Data Systems (INFOODS) has developed guidelines addressing key aspects of databases, including food and component nomenclature, data checks, and sampling, as well as providing a food composition database management system (Section 4.5) (Food and Agriculture Organization of the United Nations, 2022). International cooperation can address challenges that may pose a barrier to between-country comparisons, as well as facilitating the synthesis of findings from different studies to inform policies and programs. The Intake Center for Dietary Assessment also provides guidance on compiling food composition databases for large-scale 24h recall surveys conducted in low- and middle-income countries (Vossenaar et al., 2020).

Duplicate coding of recalls or records by independent coders has been used as a quality control procedure. Guan et al. (Guan et al., 2017) outlined a systematic method to examine discrepancies in coding, applying manual verification of codes applied to a 1% random sample of diet histories to develop a classification system outlining the types of discrepancies encountered. Discrepancy types included minor derivations not affecting food and nutrient intake estimation, incorrect codes affecting food and nutrient intake estimation, missing codes, and codes applied without corresponding foods and beverages reported in the diet histories. The classification system was then applied to a random sample of diet histories, revealing that discrepancies were most common for vegetables, followed by meats, sauces and condiments, and cereals. In one case, four slices of cheese were recorded in the diet history but coded as four cups. There were also errors in coding the frequency of consumption over the period queried by the diet history. The authors note the need to tailor the data entry and quality control procedures to the nature of the data being collected (Guan et al., 2017).

Errors in coding foods have been reduced by automating and integrating the data collection and coding processes and by allowing food codes to be generated automatically by the coder selecting the food item from a computer-based pull-down menu. Both these strategies were used for identifying and describing foods in the EPIC-SOFT program used for the EPIC study, in an effort to minimize subjective interpretations and coding error. First, a predefined country-specific list of foods, recipes, and dietary supplements was compiled (Crispim et al., 2014). These lists also included "nonspecified" generic foods which were used when respondents were not able to describe the foods adequately. Use of these generic foods minimized arbitrary decisions by the interviewers. Second, the level of detail for describing each of the foods and mixed dishes was standardized across countries (Slimani et al., 2000; Crispim et al., 2014).

A combination of automated and manual food coding is used by the U.S. Department of Agriculture to code intake data collected in NHANES(Anderson & Steinfeldt, 2004; Steinfeldt et al., 2013b; Moshfegh et al., 2022). The food and portion options included in the food composition database, the Food and Nutrient Database for Dietary Surveys, are based on prompts within the AMPM, the multiple-pass method used to carry out the 24h recalls. Within AMPM, automated coding is completed based on a matching process between the responses in AMPM and an autocode pathways database (Anderson & Steinfeldt, 2004). Table 5.3 shows three examples of autocode pathways for coffee.

Table 5.3 Examples of food paths and food code links for the coffee food category.

Food category		Sequence #	Question variable	Response value	Foodcode link
20010	1	1 2 3	CoffeeKind CoffeeCaffeine CoffeeForm	Coffee Regular Brewed	92101000 Coffee made from ground, regular
20010	2	1 2 3	CoffeeKind CoffeeCaffeine CoffeeForm	Coffee Regular Instant	92103000 Coffee made from powdered instant, regular
20010	3	1 2 3	CoffeeKind CoffeeCaffeine CoffeeForm	Coffee Regular Drink	92100500 Coffee, regular, NS as to ground or instant

Autocoding was introduced for the NHANES 2004, with automatic coding of about 45% of foods and amounts reported. For the 2017-2018 survey, the rate of automatic coding was about 60% (Moshfegh et al., 2022). Foods and amounts that are not exact matches to autocode pathways are manually matched to a food code by trained staff using SurveyNet, with automated lookup procedures to ease the process. Manual coding requires about one hour per dietary recall (Moshfegh et al., 2022). ASA24, which is based on an adaptation of the AMPM, incorporates fully automatic coding (Zimmerman et al., 2009; Subar et al., 2012), as do other web-based systems discussed in Chapter 3. Open-ended responses, entered when a respondent is not able to find the food or beverage for which they are searching or to specify details about a food reported within ASA24, are coded using "best guess" default codes based on follow-up questions to obtain general information about the item (Zimmerman et al., 2015). An examination of the implications of reviewing these codes and correcting them when necessary, found that this process may not be required for large studies, though may be worthwhile for smaller studies in which a small number of errors may be more impactful for estimated intake (Zimmerman et al., 2015).

Systems can be designed to check for outliers arising from errors in entering portions, as well as for missing information, while the respondent is still present or prior to data analysis. For example, in EPIC-SOFT, energy and macronutrient intakes from the 24h recall were calculated immediately after the interview. Calculated values were then checked against standard requirements based on the respondent's age, sex, weight, and height. Such a strategy limits a posteriori arbitrary decisions on outlier values or unlikely food data (Slimani et al., 2000). Systems like ASA24 provide researchers with information on the number of foods reported and total energy intake for a given recall to support data quality appraisal. In addition to potentially reviewing open-text entries, researchers using ASA24 are encouraged to review their data for outliers to identify possible data entry or conversion errors prior to analysis (National Cancer Institute, 2024b).

Considerations in coding eating occasions and foods consumed in combination

Discrepancies in coding meals or eating occasions can also occur and can be avoided by standardizing the codes prior to the study, and by adhering to their assignment. In the EPIC-SOFT program, the 24h recall period was divided into 11 common food consumption occasions during the day from "breakfast" to "after dinner" and during the night. A checklist was also devised and was adapted to the local dietary habits of each participating country, so that the interviewer could ensure that no major component of the food consumption occasion was forgotten (Slimani et al., 2000; Crispim et al., 2014).

To date, approaches to consider eating occasions are varied. Leech et al. (2015a) conducted a narrative review and found that most studies defined eating occasions based on how they were identified by the participant. Using data from the 2011-2012 Australian National Nutrition and Physical Activity Survey, Leech et al. (2015b) also compared the frequency of and energy intake from eating occasions, based on whether they were participant identified, based on time of day, or based on time intervals and/or energy criterion. The authors found that the different definitions affect how eating patterns are characterized and suggested that consensus on a standard definition of eating occasions is needed (Leech et al., 2015b),

In some cases, it may be beneficial to consider how foods are consumed in combination, as presenting the foods as single items only risks losing valuable information (Faber et al., 2013). Using data from the Irish National Adult Nutrition Survey, Woolhead et al. (2015) proposed a coding approach to identify common food group combinations at each eating occasion to create generic meals, such as cereal, milk, bread, and juice at breakfast. Mean nutrient compositions were calculated for each generic meal and used to conduct meal pattern analysis, with the authors noting that this approach enabled consideration of meal-based trends that were not apparent from the overall food intake data (Woolhead et al., 2015). In contrast, in assessing how individuals typically consume foods, Mason et al., (2015) applied individual food codes to each individual item, as per AMPM, and then recoded items consumed together, such as coffee with milk or the multiple ingredients in a sandwich, using combination codes.

In both cases, codes were aggregated into food groups based on the main food component. The difference in coding led to variation in the frequency of consumption of food groups and their rank order, with differences by racial identity also varying depending on the coding approach. Considering combination versus individual codes also resulted in differences in how snacks were characterized (Kuczmarski et al., 2015). The authors suggest that considering foods as they are consumed, including in combination, is important in research aiming to characterize dietary patterns (Mason et al., 2015) and that focusing on how foods are consumed in practice may inform more specific recommendations and targeted programs to improve intake (Kuczmarski et al., 2015).

Considerations in the handling of mixed dishes

Errors may also come about in coding mixed dishes. First, error may occur during when mixed dishes are broken into raw ingredients and converted to an "as consumed" form. The conversion usually involves applying adjustment factors for changes in weight due to cooking and for nutrient retention. These adjustment factors are generally derived from the literature and are discussed in Chapter 4. After applying the appropriate adjustments, the "as consumed" form is used, along with the estimate of the quantity of the mixed dish consumed by the respondent.

The assignment of the mixed dish to an appropriate food group may also result in errors. To assess alignment of dietary intake with food group recommendations, mixed dishes should be broken down, or disaggregated, into simple ingredients (i.e., single foods), which can then be classified. Within this approach, it is necessary to define systematically which items should be classified as prepared foods per se (i.e., bread, biscuits, soup, drinks) and not broken down into ingredients (Slimani et al., 2000). Assigning mixed dishes to food groups based on the primary ingredient may yield incorrect estimates for the contribution of major food groups to energy and nutrient intakes. For example, the fat and oil group may more than double its energy contribution from 5% to 12% when mixed dishes are incorrectly assigned in this way (Krebs-Smith et al., 2000). Faber et al. (2013) provide a discussion of considerations related to grouping foods.

This page titled 4.2: Sources of measurement error is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

4.3: Implications of measurement error for estimated dietary intakes

5.3.1 Misestimation of energy intakes

Misestimation of energy intake has been a focus of the dietary assessment literature. In addition to the emphasis on weight and weight change within the field more broadly, this focus is likely because error in estimation of energy is relatively large relative to error in other dietary components(Freedman et al., 2014). Almost all foods and beverages contain energy. Therefore, even small errors in the quantification of each food and beverage compound to impact overall estimates of energy intake (Subar et al., 2015). It has been recommended that researchers not rely upon self-reported food consumption data to estimate energy intake (National Cancer Institute, 2015; Subar et al., 2015). Nonetheless, daily energy intake is often considered as a proxy for the overall quality of data from a self-reported dietary assessment method, and even when interest is in other dietary components, controlling for estimated energy intake may be useful depending on the research question (Subar et al., 2015).

Individuals do not typically report energy intake; rather, they report foods and beverages consumed and the corresponding energy intake is estimated-or misestimated-using food composition databases (Chapter 4). However, the terminology energy under-reporting has often been used in the literature. The existence of total underreporting of energy intake has been documented using several different techniques (Chapter 7), including in national surveys in several countries (Beaton et al., 1997; Briefel et al., 1997; Price et al., 1997; Heerstrass et al., 1998; Zhang et al., 2000; Johansson et al., 2001; Rennie et al., 2005, 2007; Rangan et al., 2011; Murakami & Livingstone, 2015, 2016; Murakami et al., 2016, 2018; Garriguet, 2018). Underestimation of usual energy intakes may come about both due to underrecording and undereating. Underrecording is a failure to record all the items consumed during the study period or underestimating their amounts. It is a discrepancy between reported energy intake and measured energy expenditure without any change in body mass. Undereating occurs when respondents eat less than usual or less than required to maintain body weight and is accompanied by a decline in body mass (Goris & Westerterp, 1999). Although generally not as prevalent as underestimation, overestimation of energy intake also occurs. The different impacts of inaccurate reporting should both be considered in interpreting study findings, but there is often more attention to underestimation.

Burrows et al. (2019) conducted a review of research examining dietary assessment methods relative to doubly labeled water, which can be used to estimate energy intake over a 14-day period if individuals are in energy balance (Chapter 7). Based on 59 studies with adults, the authors found more frequent misestimation among females than males. The degree of underreporting was variable across studies, with less variation and degree of underestimation of energy intake in studies using 24h recalls (Burrows et al., 2019). Studies also suggest that misreporting is associated with weight status, socioeconomic status, and sex (Chapter 7).

Efforts to overcome the problem of energy misestimation have led some investigators to exclude presumed underreporters from the data set, using techniques discussed in Chapter 7. Such an approach introduces a source of unknown bias into the data set and is not recommended. Further, energy is only one dietary component and other dietary components of interest, including nutrient densities, are not subject to the same level of error(Freedman et al., 2014; 2015). Therefore, corrections for low energy intake are not sufficient to eliminate the biases arising from selective misreporting of certain food types (e.g., foods of low social desirability). Others advocate the inclusion of data from all respondents, but the use of statistical methods that control for energy intake. Several approaches to energy adjustment exist (Kipnis et al., 1993, 1997; Kohlmeier & Bellach et al., 1995; Mackerras, 1996; Willett et al., 1997; Hu et al., 1999; Tomova et al., 2022). The selection of an appropriate model for energy adjustment depends on the research question; readers are advised to consult a statistician on this issue.

Prior studies have suggested that certain foods, such as cakes and pies and savoury snacks, may be underestimated to a greater extent than others(Poppitt et al., 1998; Krebs-Smith et al., 2000). In a systematic review of 29 studies, Whitton et al. (2002) concluded that there was more variation in error within food groups and within studies than between food groups. Most studies examined used 24h recalls and controlled feeding designs. Beverages appeared to be omitted less frequently than other items, but estimation of their portion sizes was less accurate. Foods like desserts and snacks did not appear to be omitted more frequently than foods like fruits and vegetables. Rather, errors in estimation appeared to be related to their form, e.g., meat in slices or as single units and vegetables and condiments as part of mixed dishes, as discussed in Section 5.2.1. The authors suggested that omissions tend to be items that are less visible to participants rather than those that they perceive as less healthy (Whitton et al., 2002).

5.3.2 Portion size misestimation

The errors associated with quantifying the portion of food consumed are probably the largest measurement error in most dietary assessment methods. They can arise from the respondent's inability to quantify accurately the amount of food consumed, or from misconceptions of an "average" portion size.

SHARON MAY UPDATE HERE IN RELATION TO Chapter 7 Respondents differ in their ability to accurately estimate portion sizes visually. In general, such discrepancies appear to be independent of age, body weight, social status, and gender of the respondent, but they do vary with the type and size of food (Young & Nestle, 1995). Large errors may occur, for example, for estimates of foods high in volume but low in weight (Gittelsohn et al., 1994) and for intact cuts of meat of irregular shape (Godwin et al., 2001). Furthermore, respondents appear to have greater difficulty estimating the size of large portions than they do small portions, irrespective of their body weight (Young & Nestle, 1995).

Due to the considerable challenge that portion size estimation presents for accurately estimating dietary intake, it continues to be an area of innovation within the field. Several types of portion-size measurement aids (PSMAs) have been developed in an effort to enhance the accuracy of portion size estimates when weighing methods cannot be used. Amoutzopoulos et al. conducted a systematic review of PSMAs (2020), identifying over 500 aids used in over 300nbsp;studies. The authors found that the measurement aids most often used were 2- and 3-dimensional aids, mainly household utensils and photographic atlases, whereas 1-dimensional aids, such as portion lists and food guides, were used less often. Most of the aids assessed were developed and/or used in high-income countries, mainly the U.S. and United Kingdom (Amoutzopoulos et al., 2020). However, food images, including food atlases, can be useful for low- and middle-income settings with tailoring of the images to foods consumed in the region. For example, adaptation of GloboDiet for use in different regions requires consideration of the suitability of the photos, as well as the household measures and standard units, used in the region (Aglago et al., 2017). In low-income countries, real foods, food models, or salted replicas of actual staple foods may be useful (Ferguson et al., 1995; Gibson & Ferguson, 2008; Aglago et al., 2017). In all cases, measurement aids that depict a range of portion sizes should be used to avoid the tendency to "direct" responses.

Photographs and digital images are being increasingly used to assist respondents in estimating portion sizes; several approaches have been used and are discussed in Section 3.2.2. Photographs that depict a range of portion sizes are often used to aid in the estimation of portion size in 24h recalls and food frequency questionnaires. Photographs and images for quantifying portion sizes should be standardized with respect to several factors. Graduated food models and household measures have been used in several national food consumption surveys. USDA uses three-dimensional measuring guides for the first recall, completed in person, and the USDA Food Model Booklet (Ingwersen et al., 2007; Centers for Disease Control, 2015), which contains two-dimensional drawings that are consistent with the measuring guides, for the second recall (Moshfegh et al., 2022). Cleveland and Ingwersen (2001) compared the portion-size reporting accuracy of two-dimensional (2D) food models and a range of 3D measurement aids used in the dietary component of NHANES. The 2D food models include 32 life-size drawings of household vessels (glasses, mugs, bowls), abstract shapes (mounds and spreads), and geometric models (circles, a grid, wedges, and thickness bars). The 3D measurement aids are actual measuring cups, spoons, and a ruler. Overall, both the 2D and the 3D guides helped generate relatively good estimates of food amounts, although in this study, more accurate estimates were obtained on average with the 2D than the 3D guides, especially for mounded foods.

SHARON MAY UPDATE HERE IN RELATION TO Chapter 7 Some studies have investigated whether it is helpful to train respondents to use portion-size measurement aids such as graduated food models or household measures. In general, the use of short group training sessions for respondents using food models or household measures should be encouraged. Training sessions enhance the ability of both children and adults to estimate food portion sizes accurately, although for children, more than one training session is probably necessary (Bolland et al., 1990; Weber et al., 1999). Training using a combination of food models and life-sized food photographs may be best (Howat et al., 1994).

Quantifying standard reference portion sizes

Semiquantitative food frequency questionnaires, used to rank individuals according to food or nutrient intake, often specify a standard reference portion size for each specific food. Typically, this is intended to represent the median amount consumed during a single meal. The values may be generated from country-specific national nutrition surveys (Block et al., 1986) or other large surveys (Willett et al., 1985). Respondents appear to have difficulty relating what they consumed to such predefined standard reference portion sizes, (Friedenreich et al., 1992), and inconsistent results and large error have been reported (Willett, 1994). ().

Many factors affect actual food portion sizes, including age, gender, activity level, the appetite of the individual, household utensils

used, and where and when the food is obtained and eaten.

Guthrie (1984) compared the amount of food selected as a usual portion size by young adults, in relation to the U.S. standard reference portion sizes at that time. The latter were based on median portions derived from USDA Nationwide Food Consumption Survey data (Pao et al., 1982). (). Items such as butter on toast, sugar on cereal, milk as a beverage, and tossed salad corresponded closely with standards, whereas others such as dry cereals, orange juice, and fruit salad did not. Men tended to select larger portions than women.

In view of these findings, the use of standard reference portion sizes in food frequency questionnaires is still hotly debated. Some investigators have argued that because variations in food intake are mainly determined by frequency of consumption, obtaining information on portion sizes in semiquantitative food frequency questionnaires is not always justified (Samet et al., 1984; Noethlings et al., 2003). Others recommend the use of small, medium, and large portions, based on age and gender-specific median portion sizes as standards (Cummings et al., 1987). Japanese investigators have cautioned that the relative contributions of within- and between-person variations in portion size vary among food items, so that whether separate questions on portion size should be included will depend on the food groups relevant to the diet-disease association being studied (Tsubono et al., 1997). ().

There is some confusion in the literature between the use of standard reference portion sizes and standard reference serving sizes. Two examples of standard reference serving sizes established for use in the Food Guide Pyramid (FGP) by USDA (USDA, 1992), (). and for food labels by the U.S. Food and Drug administration (FDA, 1993), (). are shown in Table 5.3, and compared with standard reference portion sizes generated by Willett et al.(1985) and Block et al. (1986). Inconsistencies occur between the standard reference serving sizes and standard reference portion sizes.

5.3.3 Omission of information on nutrient supplement usage

The appropriate consideration of dietary supplement use in nutrition surveys conducted in high-income countries is critically important. In the United States, for example, more than half of adults and one-third of children use dietary supplements (Bailey et al., 2011b, 2019; Kantor et al., 2016; Stierman et al., 2020). In Australia, it is estimated that close to half of women and over one-third of men use dietary supplements (Burnett et al., 2017). Not accounting for supplement use may result in a systematic underestimate of intakes of certain nutrients, and thus an overestimate of the prevalence of nutrient inadequacy (Bailey et al., 2011a; Dwyer et al., 2022). Not considering supplement usage could also result in underestimation of the prevalence of excessive intake. Inclusion of dietary supplement use may be particularly important in accurately assessing dietary intake at certain life stages, such as the prenatal and periconceptional periods (Dwyer et al., 2022). Assessing supplement usage requires unique considerations compared to food because supplements may be consumed daily or episodically, with usage varying substantially over time, and can provide high levels of nutrients that are not associated with energy intake (Bailey et al., 2019).

Supplement use has been assessed using frequency-type questionnaires, inventories, short screeners, and in conjunction with assessments of food intake using 24h recalls and food records. Information on the validity, reproducibility, and measurement error structure of methods to assess supplement use is limited (Bailey et al., 2019). Discrepancies exist in the terminology and the methods used to measure dietary supplements and the criteria used to define dietary supplement users, thus limiting comparisons across studies (Brownie & Myers, 2004). For example, considerable variability has been identified in how supplement use is queried by widely used food frequency questionnaires (Rios-Avila et al., 2017).

The U.K. National Diet and Nutrition Survey of people aged 65y or over obtained information on supplement usage from responses on a health and lifestyle questionnaire and a 7d weighed diet record. A comparison of the total nutrient intakes and the corresponding blood biochemical indices suggested that the 7d record was not long enough to record habitual dietary supplement use. Instead, a structured questionnaire probing for supplement use over a longer period was recommended. This included close-ended questions about the specific brand taken, the amount per pill, the frequency of use, and the duration of use (Bates et al., 1998, 2000). A similar conclusion was reached by Patterson et al. (1998a). These investigators also emphasized that measurement error associated with long-term supplemental vitamin and mineral intake may be responsible for the lack of any observed association between vitamin supplements and the risk of cancer. Since 2007, NHANES has included the collection of dietary supplement information at the end of the 24h recalls, in addition to an inventory and in-home questionnaire that collects information on amounts typically taken, frequency of use in the past 30 days, and motivation for taking the supplement (Gahche et al., 2018; Bailey et al., 2019). Usage estimates are lower based on the 24h recall versus the 30d questionnaire. Nicastro et al. thus recommended using information from both (2015).

Accurate information on brand names is critical for dietary supplements because inter-brand variability is large. Failure to correctly quantify the dose of a supplement can have a greater impact on the estimation of nutrient intakes than from any source

of food intake underreporting. Additionally, the chemical form of the dietary supplements can affect their bioavailability, so it is preferable to record the chemical characteristics of the dietary supplements, whenever possible (Heimbach, 2001). This can be achieved by asking participants to have the dietary supplements that they take available. In this way, the interviewer can ensure the type and amounts recorded are correct (Patterson et al., 1998b).

Maintaining supplement databases is challenging because of the number of products on the market and their rapid turnover (Bailey et al., 2019; Dwyer et al., 2022). Dwyer et al. (2022) provide an in-depth discussion of considerations related to supplement databases. In the US, the Dietary supplement Ingredient Database (Office of Dietary supplements and US Department of Agriculture, 2023) has been developed to provide information on the nutrient values of dietary supplements. supplement databases have also been compiled in other countries (RIVM, n.d.; Food Standards Australia New Zealand, 2024). Through a joint project of the FAO, INFOODS, and the George Institute-Australia, a global dietary supplement database is under development (The George Institute for Global Health, 2024). The database will be open source and is intended to promote the consideration of supplements in estimating nutrient intake

Children

The accurate assessment of dietary intake in children is especially challenging. Children tend to have diets that are highly variable from day to day, and their food practices often change markedly across life stages. Research on the development and timing of specific cues that may help children report their diets more accurately, applying a cognitive-processing approach, has been conducted (Baranowski & Domel, 1994).

Warren et al. (2003) concluded that children aged 5-7 y were unable to provide an accurate dietary recall of their school lunch, especially when they consumed a dinner provided by the school rather than eating their own packed lunch, as shown in Figure 5.2. Nevertheless, prompts and cues enhanced recall by all children in this study. Main dishes were remembered best by the children; leftovers were not readily reported. A series of recommendations and suggestions for future studies on children have been compiled by these investigators and are shown in Box 5.2 (Warren et al., 2003). There is no doubt that more work is needed on methods to determine more accurately what children aged <8y are eating.

Baxter et al., 2015 have noted that the retention interval and level of probing, which can both be determined by the researcher depending on the recall protocol or system used, interact to influence the accuracy of reporting among children. Furthermore, the level of accuracy in relation to different combinations of retention intervals and types of prompting differed by gender.

This page titled 4.3: Implications of measurement error for estimated dietary intakes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

4.4: Minimizing measurement error through data collection procedures

Measurement error can be minimized by incorporating quality-control procedures into each stage of the dietary assessment method. These procedures include standardization of interviewing techniques and questionnaire, robust training of interviewers and coders, pretesting of questionnaires, and administration of a pilot study. Each procedure must be checked continuously to ensure compliance with standardized protocols.

The existence of measurement error continues to be a major challenge in nutrition surveillance and research and provides an ongoing impetus for innovation in dietary assessment. Technology-enabled methods, such as online recalls incorporating multiple passes and automated coding and web-based food frequency questionnaires integrating automated skip patterns, are increasingly used. It remains important to consider what is known about the validity and reproducibility of a given method for the population of interest and to integrate a pilot study to ensure the method performs as intended and to identify opportunities to reduce error.

Random error, unlike systematic error, can be minimized by increasing the number of observations. Random error may occur across all respondents and all intake days. In contrast, systematic error may be more common and/or larger among some respondents (e.g., individuals in larger bodies), in data collected by specific interviewers or using different methods, or for certain foods (e.g., alcohol). Systematic error may be mitigated by ensuring that the methods used, such as frequency questionnaires, are appropriately tailored to the population group(s) of interest, as well as by calibrating data to a more accurate method administered in at least a subsample.

As noted, measurement error can be accounted for or mitigated if repeat administrations (random error) or reference data from another method (systematic error) are available. It is ideal if these data are available for a random subsample of the overall sample of interest but in some cases, estimates from an external sample with similar characteristics can be used.

This page titled 4.4: Minimizing measurement error through data collection procedures is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

4.5: Summary

This chapter outlines the systematic and random error that may occur during the collection and recording of food consumption data. In practice, different types of error compound to impact the overall accuracy of reported consumption and estimated intake of nutrients and other dietary components. For example, errors in portion size quantification may counteract or exacerbate the impact of omissions of foods and beverages consumed on nutrient intake estimates.

Quality-control procedures that minimize possible sources of measurement error include training the interviewing and coding staff and developing standard interviewing techniques and questionnaires during the pilot survey. Increasingly, sources of error arising from both respondent and interviewer biases and respondent memory lapses can be reduced using computerized probing questions, standardized prompts, and built-in cues during automated dietary interviews, as well as technology-enabled methods. Nevertheless, misestimation of energy and selective misreporting of certain food types remain important sources of respondent biases. A variety of portion-size measurement aids are now available for use when weighing methods are not possible. These include the use of 2-D graduated food models, photographs, and images and 3-D measurement guides (e.g., household measures) to quantify portions of foods consumed. Training respondents to use these measurement guides to estimate food portion sizes will also improve accuracy. Collection of accurate data on consumer use of dietary supplements is also essential; information on brand, dosage, chemical form, and period over which use of the dietary supplement has been recorded is required.

Establishing a computerized standard coding system for both foods and eating occasions to avoid coding error is critical, especially for surveillance and cross-country comparisons. Systematic detection of wrongly coded weights of foods is more difficult, although calculation of energy and macronutrient intakes from 24h recall interviews, while the respondent is still present, allows the correction of any gross error. Finally, care must be taken in the handling of data for mixed dishes and foods eaten in combination. Despite all efforts to minimize sources of random and systematic error that may occur during the measurement of food and nutrient intakes, some errors remain difficult to predict and to prevent and, as a result, may introduce differential bias in reported food intakes. Ongoing research is needed to investigate the specific type and nature of measurement error, especially in diverse populations, so that these can be minimized or corrected statistically. In this way, the analysis and interpretation of dietary data can be improved. The existence of dietary measurement error distorts estimates of disease relative risk, and thus has major implications for epidemiological studies of dietary risk factors and disease. Observed diet-disease relationships should be interpreted cautiously (Freedman et al., 2011), particularly if steps have not been taken to mitigate the error to the extent possible.

Assessing the reproducibility (Chapter 6) and validity (Chapter 7) of dietary methods used is essential for surveillance, including cross-country comparisons, epidemiologic, and intervention research (Buzzard & Sievert, 1994).

This page titled 4.5: Summary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5: NUTRIENT REFERENCE VALUES (CHAPTER 8A)

Abstract

This chapter describes the evolving process of setting nutrient reference values (NRVs) from a single value to a multi-level approach designed to address multiple users and needs. This multi-level approach with some modifications has been adopted by several authoritative groups. However, discrepancies in the terminology and methodologies has led to an initiative to develop a four-step framework for harmonizing the process for deriving two core NRVs. The feasibility of the proposed framework has been tested with three exemplar micronutrients (zinc, iron, folate) on two high-risk population sub-groups — young children and women of reproductive age. The framework can be applied across countries and a variety of population subgroups, while taking into account culturally and context-specific food choices and dietary patterns. Derivation of the two core NRVs depicted in the framework — the Average Requirement (AR) and Safe Upper Level of Intake (UL) — is presented here. Details of additional NRVs used are also described. They include the Recommended Intake (RI) (analogous to the RDA or RNI) used to guide intakes of individuals, the Lower Reference Intake (equivalent to the AR –2SD), and a Safe or Adequate Intake value used when insufficient information is available to set an AR. Details of the different methods used to establish recommendations for energy and macronutrients are also described. A review of the NRVs currently in use by the United Kingdom, USA and Canada, European Union, and WHO/FAO is also presented, together with the potential sources of discrepancy. In a final section under "Future directions", reasons for harmonizing the process to derive NRVs globally are emphasized together with the six core values considered essential for the success of this initiative.

5.1: Development of Nutrient Reference Values (8a.1)

5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)

5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)

5.4: U.K. Dietary Reference Values (8a.4)

5.5: United States and Canada- Dietary Reference Intakes (8a.5)

5.6: European Dietary Reference Values (8a.6)

5.7: WHO/FAO Nutrient Reference Values (8a.7)

5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)

This page titled 5: Nutrient reference values (Chapter 8a) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Development of Nutrient Reference Values (8a.1)

The first set of nutrient reference values (NRVs) was produced by the Technical Commission on Nutrition, League of Nations, (1938). The recommendations of this Commission formed the basis for the first Canadian Dietary Standard compiled by the Canadian Council on Nutrition (1940). The United States Food and Nutrition Board (in 1943) prepared the first U.S. Recommended Dietary Allowances (RDAs) for protein, energy, and eight vitamins and minerals. This was followed by periodic revisions of the U.S RDAs and a number of joint publications by the World Health Organization / Food and Agriculture Organization on recommendations for different groups of nutrients (FAO/WHO, 1974; FAO/WHO, 1988). By 1989, the U.S RDAs included 25 vitamins and minerals in addition to protein and energy.

Until this time, these reports provided a single estimate of requirements (i.e., the RDA or equivalent) that was sufficient to meet the needs of almost all individuals in a specific life-stage group. Nevertheless, despite being set for individuals, the RDAs were being applied to population data and used for many different purposes. This misuse of the RDA led to the development of new approaches to address multiple users and needs. The first country to adopt a framework that embraced three nutrient reference levels was the United Kingdom(COMA,1991) (Section 8a.4). This was followed by the release by the U.S. Food and Nutrition Board (FNB) of a new paradigm of Dietary Reference Intakes (DRIs)(National Academic Press, 2000). Their approach, developed jointly by the U.S. and Canada, also provided multiple reference levels for each nutrient and included for the first time an upper intake value (UL) as well as the concept of reducing the risk of chronic diseases

This new multi-level approach was soon adopted with some modifications by other authoritative bodies worldwide, including in China; Korea and Southeast Asia; Germany, Austria, Switzerland; Australia and New Zealand; and Mexico. Most established an Estimated Average Requirement (EAR) and a RDA (or equivalent) at two standard deviations above the EAR, although the exact terms used differed. Some authorities also included a Tolerable Upper Intake Level (UL) for certain nutrients, and in many cases, a safe or adequate intake value (or range) when insufficient information was available to set an EAR (King et al., 2007).

One of the latest developments includes the introduction of a systematic review process to identify data relevant to the derivation of NRVs. This initiative was introduced in 2007 by the EURopean micronutrient RECommendations Aligned (EURRECA) Network of Excellence (NoE) (Dhonukshe-Rutten et al., 2013), and used in the U.S. and Canada for the first time for the revision of the NRVs for vitamin D and calcium (IOM, 2011). More recently, efforts have been made to overcome some of the challenges encountered when setting nutrient reference values for chronic disease prevention; these efforts are on-going (Yetley et al., 2017).

There is now considerable interest in providing a framework for harmonizing the process to derive NRVs that can be applied across countries and a variety of population subgroups, while taking into account culturally and context-specific food choices and dietary patterns. This international harmonizing initiative first began in 2005 with the development and publication of several background review papers and a consensus report on “Harmonization of nutrient intake values” (King and Garza, 2007). The first harmonization initiative was followed a decade later with an international workshop convened in 2017 by the National Academies of Sciences, Engineering, and Medicine (NASEM) in partnership with WHO and FAO to explore the evidence for achieving global harmonization of methodological approaches to derive NIVs across countries; see NASEM (2018a) for the report of the workshop proceedings. (Table 8a.1). At this workshop, the term Nutrient Reference Values (NRVs) was adopted to describe collectively the nutrient intake recommendations and is analogous to terms currently in use such as Dietary Reference Intakes (DRIs) in the United States and Canada and Dietary Reference Values (DRVs) in the United Kingdom and the European Union.

Table 8a.1: Comparison of the suggested “recommended” terminology with terms in use at present by USA/Canada, UK, European U AI, adequate intake; AMDR, adequate macronutrient distribution range; AR, average requirement; DRI, dietary reference intake; DRV, LTI, lower threshold intake; NIV, nutrient intake value; PRI, population reference intake; RDA, recommended dietary allowance; RI, re intake (WHO/FAO); SIV, Safe intake value; SUL, Safe upper level; UL, tolerable upper intake level (USA/Canada & EFSA), upper tolera Adapted from King and Garza (2007).

Recommendation	Recommended terms	USA/Canada
Umbrella term for the set of recommendations	NRV	DRI
Average requirement	AR	EAR
Recommended intake level	RI	RDA
Lower reference intake		

Recommendation	Recommended terms	USA/Canada
Adequate intake	AI	AI
Safe upper level of intake	UL	UL
Appropriate macronutrient distribution range		AMDR

Recommendations from the workshop for the terms for three reference values and the relationship between them is shown in Figure 8a.1. The three reference values are: the Average Requirement (AR), Recommended Intake (RI), and Safe Upper intake level (UL). Of these, both the AR and UL are used to plan, assess, and evaluate nutrient intakes at the population level, whereas the RI is the appropriate reference to use for the assessment and planning of intakes for an individual. However, when there is insufficient evidence to set an AR, then an Adequate Intake (AI) is derived based on observed or experimentally determined estimates for an apparently healthy population. For more details of the derivation of these core reference values, see Sections 8a.2.4 for AR and Section 8a.3 for RI and the AI. This initial organizing framework emphasizes on the left the concepts that serve as the basis for setting the two recommended NRVs. Their uses at both the individual and population level are shown on the right, along with other critical health applications of the NRVs. The two core reference values — the AR and the UL — are prioritized, from which other reference values are derived. The individual recommended intake level (RI) is used to guide intakes at the individual level, and is conventionally set to cover the needs of 98% of individuals.

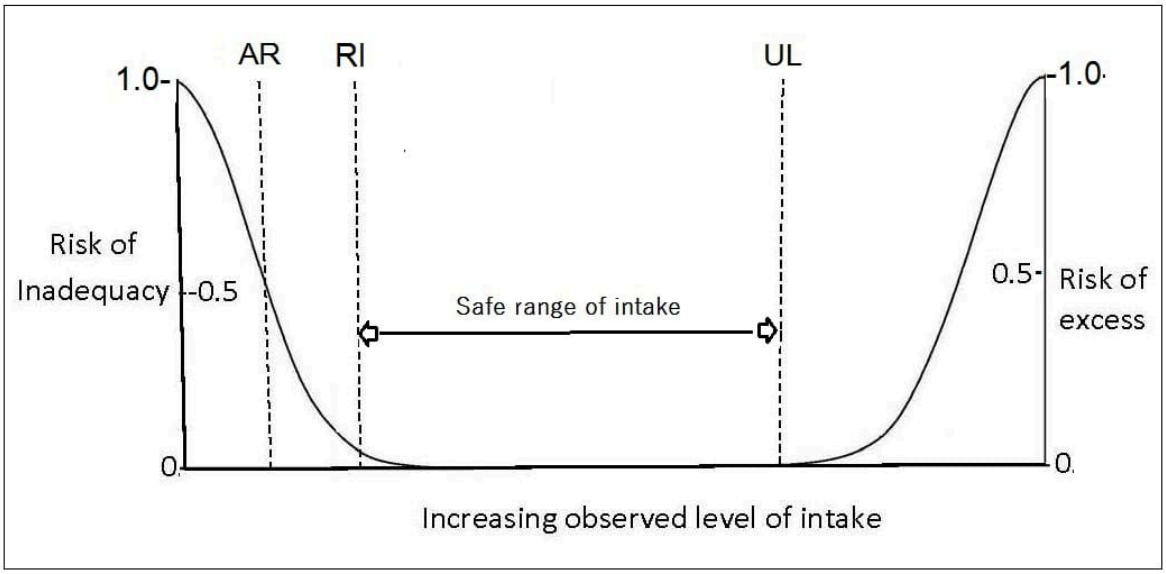


Figure 8a.1: Risk of inadequacy (left y-axis) and risk of excess (right y-axis) as a function of observed level of intake (x-axis). Modified from NASEM (2018).

This page titled 5.1: Development of Nutrient Reference Values (8a.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)

An initial organizing framework for deriving NRVs, developed in the 2005 workshop(King and Garza, 2007), is depicted inFigure 8a.2.

>

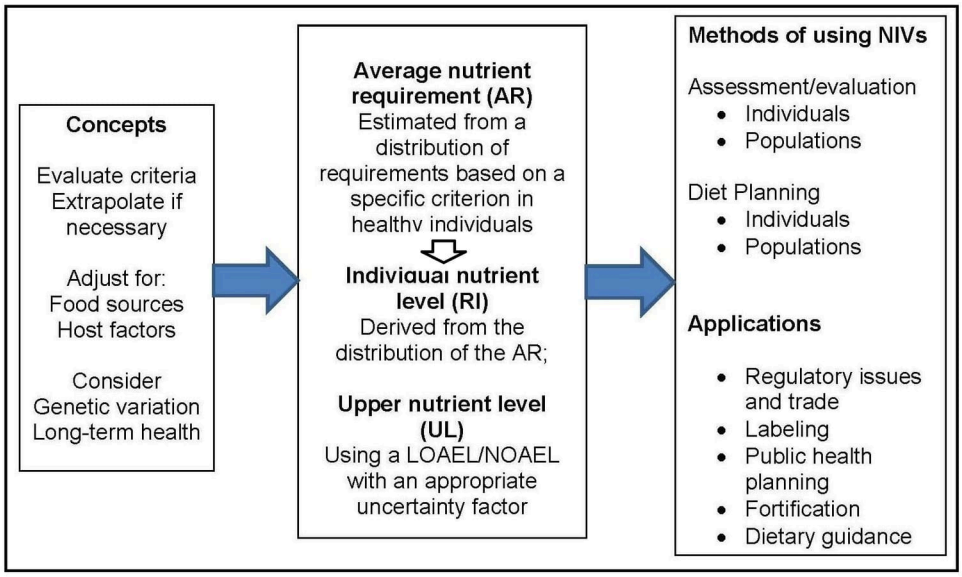


Figure 8a.2: An initial organizing framework for deriving NRVs. Adapted from King and Garza (2007).

Of the two core reference values, the AR is defined as the intake needed by 50% of a population subgroup (based on age, gender, and physiological status) to meet a specific criterion of nutrient adequacy. NASEM (2018) emphasize that the AR should:

- Be based on the mean nutrient intake of a specific population;
- Be established for all essential nutrients and food components that have public health relevance;
- Include acceptable macronutrient distribution ranges for carbohydrates, protein, and fat that reduce chronic disease risks associated with the intake of these macronutrients;
- Consider nutrient-nutrient interactions* and quantify them, if possible; and
- Consider subpopulations with special needs, keeping in mind, however, that reference values are intended for apparently healthy** individuals.

* Some nutrient-nutrient interactions are now known to alter requirements (e.g., calcium-protein-sodium, protein-energy; vitamin E and polyunsaturated fatty acids).

** The term “apparently healthy” has come into question because of the implication of rising global prevalence of overweight and obesity and the corresponding increase in chronic disease risk. The NASEM (2017) report on developing dietary reference intakes based on chronic disease recommends that for future NRVs, the health status of the population should be characterized in terms of who is included and excluded for each NRV.

The second core reference value — the UL — is defined as the highest level of usual daily nutrient intake that poses no risk of adverse health effects in most individuals in the population and is based on a toxicological risk assessment model (IOM, 1998). For details of the derivation of the AR and UL, see Sections 8a.2.4 and 8a.2.5.

Based on the deliberations of a second workshop held in 2017 that reviewed the strengths and weaknesses of the methods

currently in use, a new framework for harmonizing the process for deriving the two core nutrient reference values (i.e., AR and UL) was developed. This is presented in Figure 8a.3 that shows the four major steps required to estimate the key NRVs, with the components needed to complete the steps itemized. The four steps are:

- Choosing the appropriate tools and resources;
- Collecting relevant data from the tools and other resources;
- Identifying the best approach for the nutrient under consideration; and
- Deriving the two core nutrient reference values, the average requirement (AR) and the tolerable Upper Limit (UL).

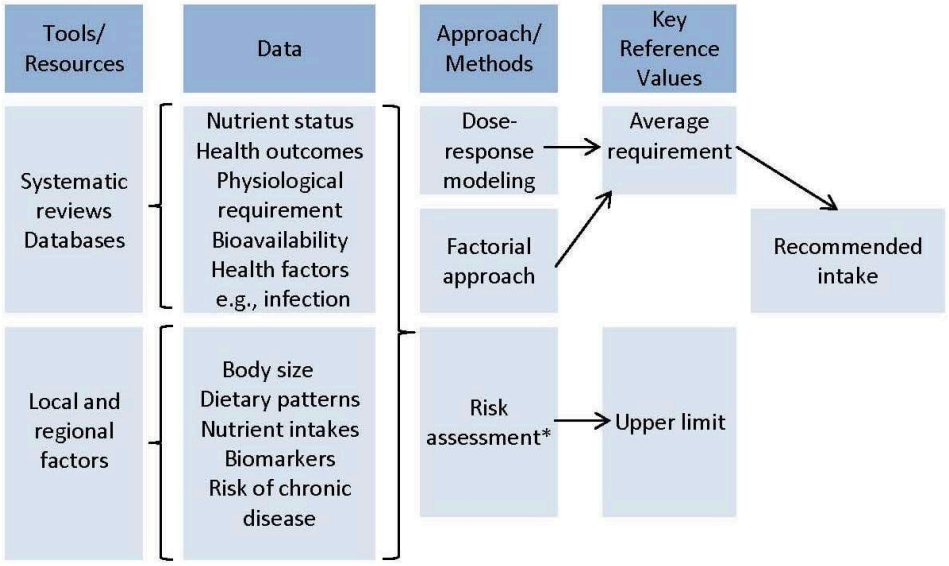


Figure 8a.3: Framework for harmonizing the process to derive NRVs. Modified from Russell et al. (2018b).

The feasibility of this proposed harmonization framework was tested using three exemplar nutrients — zinc, iron, and folate, nutrients of concern for young children and women of reproductive age; for more details, see NASEM (2018b).

8a.2.1 Choosing the appropriate tools and resources

The primary tools and resources needed to develop the NRVs are: systematic reviews, comprehensive databases, and information about relevant local and regional factors that can influence the NRVs, as shown in Figure 8a.3. The EURECCA network applied the systematic review process to identify data relevant to the derivation of NRVs, using six micronutrients as examples; see Hooper et al., 2009; Van 't Veer et al., 2013; and Dhonukshe-Rutten et al., 2013 for more details. Subsequently, systematic reviews of biomarkers of status for vitamin B_{12}, zinc, iodine, copper, riboflavin, magnesium, vitamin D, polyphenols, n-3 long-chain poly-unsaturated fatty acids, and selenium were published (Allen et al., 2009; Pérez-Jiménez et al., 2010; Witkowski et al., 2011). These existing systematic reviews can be updated, or new systematic reviews initiated, where necessary.

Efforts have been made to harmonize the protocols for systematic reviews (Moher and Tricco, 2008). The first step is to use the PICO/PECO model to define the search terms and concepts for the systematic review; this model is described in Box 8a.1.

Box 8a.1: The PICO/PECO Model

The elements in the model are:

- P = Population: How would you describe the population subgroup? What are the most important characteristics of the population?
- I/E = Intervention/Exposure (I/E): What primary intervention or exposure are you considering?
- C = Comparison: What is the main alternative to compare with the intervention?

- O = Outcome: What is the outcome or effect being considered?

Modified from NASEM (2018b)

In addition to using the PICO/PECO model, construction of a predefined analytic framework (e.g., a causal pathway) is also recommended to help identify systematic review questions. An example for a generic analytic framework for NRVs is depicted in Figure 8a.4.

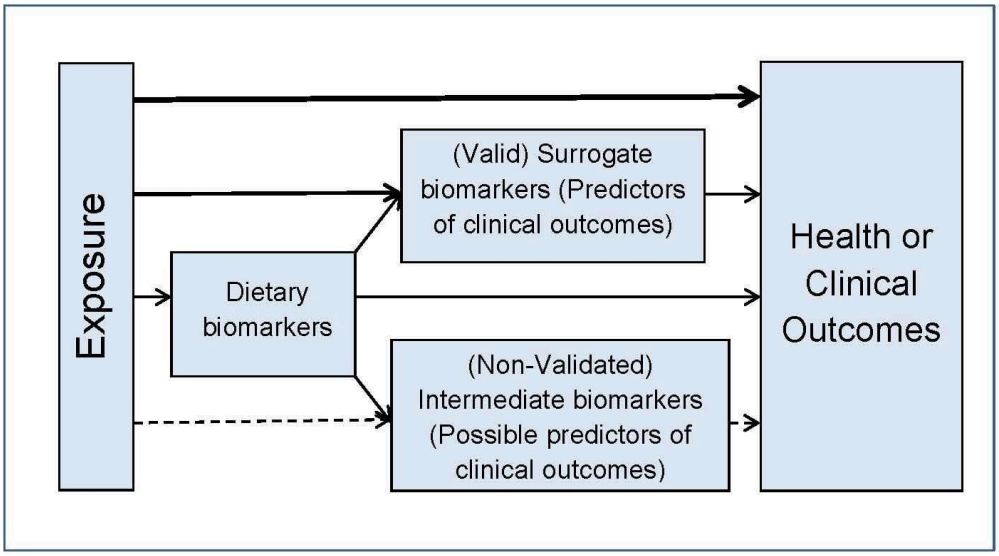


Figure 8a.4: Generic analytic framework for a systematic review of studies on the association between a nutrient and health outcomes. Modified from Russell et al. (2009) and NASEM (2018b).

The representation includes putative associations between an exposure (e.g. a nutrient) and dietary biomarkers of intake (e.g., status biomarkers such as serum or tissue nutrient concentrations, (non-validated) intermediate biomarkers (possible predictors of health or clinical outcomes), (valid) surrogate biomarkers (predictors of health or clinical outcomes), and health or clinical outcomes. The solid arrows represent established associations among factors. Line thickness represents the relative directness of an association and the strength of the relation with the health or clinical outcome. Dotted lines represent associations to surrogate biomarkers for which there is no good evidence of an association. Surrogate biomarkers are often used when the study duration is too short to show an effect on health or clinical outcome.

The analytic framework describes the relationships between “exposure” (i.e., nutrient intake) and outcomes of interest, and helps to emphasize what aspects are known and unknown. Note that the analytic framework should be modified to reflect the underlying biological factors associated with a specific nutrient, the life-stage group, and the reference value of interest (i.e., AR or UL). For example, if the AR is the reference value of interest, then the outcome could be a clinical or health condition or a surrogate biomarker (preferably a functional biomarker of nutrient status) associated with deficiency of the nutrient, whereas for the UL, the clinical or health condition or surrogate biomarker associated with nutrient excess would be selected; see Hooper et al. (2009) and Calder et al. (2017) for details on the selection of reliable surrogate biomarkers. In most cases, a single outcome that serves as a measure of exposure is selected for the AR or UL for a specific nutrient, sex, and life-stage group.

The review of the literature for the systematic review should be guided by the completed PICO elements, associated inclusion criteria, and the analytic framework for the nutrient, life-stage group, and NRV of interest. Quality assessment instruments (QAIs) must be used to evaluate the evidence at the individual study level, the choice depending on the study design. Numerous QAIs are available. SIGN 50 is an example of a QAI that can be used for both randomized controlled trials and observational studies. Alternatively RoB 2, a revised Cochrane risk-of-bias tool for randomized trials, can be used (Sterne et al., 2019).

The strength of the body of evidence generated from a systematic review must also be evaluated using QAIs. Examples include PRISMA , AMSTAR2 (Shea et al., 2017), and GRADE (Balshem et al., 2011). The latter has been used by both Australia/New Zealand and WHO in their revisions of Nutrient Reference Values. These tools have quality descriptors for criteria such as study selection

and data exclusion, risk of bias, sources of funding, inclusion of nonrandomized studies, and issues around heterogeneity, etc. Application of QAIs at both the individual study level and systematic review level serves to enhance the scientific rigor and transparency of the decision- making process for deriving the NRVs. Note: nutrition-specific QAIs are under development and will be available in the future.

Readers interested in conducting a systematic review are advised to consult the Systematic Review Data Repository (SRDR). This is an open access, Web-based repository of systematic review data compiled by the U.S. Agency for Healthcare Research and Quality and is available free to users worldwide.

8a.2.2 Collecting data from the tools

The second step is to collect the data generated from the tools that are essential for selecting the biomarkers of status, surrogate outcomes, and health/clinical outcomes. Data on the dietary factors with the potential to influence nutrient bioavailability and the health factors (e.g., infection) that can affect nutrient requirements must also be included.

8a.2.3 Identifying the best approach for deriving the NRVs for the nutrient of interest

Once the relevant data have been collated, the evidence appraised and integrated using the appropriate resources, and any sources of uncertainty identified, the third step is to identify the best approach for deriving the NRVs for the nutrient under study. The decision on which approach to use depends on the availability of data, and the types and quality of studies reviewed (Section 8a.2.1). Three approaches are frequently used: factorial approach, balance studies, and an intake (dose)-response assessment. Limitations of each of these approaches were noted by Claessens et al.(2013). Table 8a.2 provides examples of the types of studies used in the development of the U.S / Canadian Dietary Reference Intakes.

Table 8a.2: Types of studies and examples of their application in the development of the U.S / Canadian Dietary Reference Intakes. Modified from Yates. (2007).

Type of study	Measurement	Examples
Nutrition intervention studies (randomized, placebo-controlled studies)	Functional outcome	Calcium fracture rate with increased calcium intake via supplements or placebo
	Biochemical measurements	Red blood cell folate response to varying levels of folate
Depletion/repletion studies	Biochemical measurements	Leukocyte ascorbate concentrations for vitamin C Urinary excretion of 4-pyridoxic acid for vitamin B_6
Balance studies	Controlled intake and excretion	Protein requirements
Factorial estimation	Measure losses + bioavailability	Iron & zinc requirements
Epidemiologic observational studies	Estimate intake and measure losses	Iodine intake and excretion
	Functional outcome	Vitamin A and night-blindness
Observed intakes in healthy populations	Dietary intake data	Vitamin K

The factorial approach is currrently used to determine NRVs for iron and zinc, as in these cases the relationship between exposure (nutrient intake) and surrogate biomarkers or a clinical/health outcome is weak or non-existent and cannot be derived mathematically. Factorial estimates are derived from the sum of obligatory losses (i.e, through fecal, urine, skin, menses etc) plus any additional needs for growth and development (fetus, pregnancy, lactation etc), adjusted by a bioavailability factor to convert

the physiological requirement into the dietary requirement (Fairweather-Tait and Collings, 2010). The Average Requirement (AR) is derived from a resultant pooled estimate of needs, taking into account the bioavailability (Dhonukshe-Rutten et al., 2013). State-of-the-art isotope techniques are now used to measure nutrient fluxes and "true" absorption and retention rates. The potential influence of diet- and host-related factors on nutrient bioavailability was reviewed by Gibson (2007).

Balance studies are used when no reliable biomarker representative of actual nutrient status exists. They measure input and excretion — when they are equal it is assumed that the body is saturated. Additional assumptions are that the size of the body pool of the nutrient is appropriate, and that increasing the levels of nutrient intake do not provide additional benefit. Limitations of balance studies have been reviewed by Mertz(2007). Balance studies are used to determine protein and, in some cases, mineral requirements (e.g., calcium and molybdenum).

Intake (dose)-response modeling, usually based on randomized controlled trials (RCTs) and epidemiological studies, describes how a known physiological outcome changes according to the exposure (i.e., intake) of a nutrient. It is used when there is a clear relationship between the exposure (i.e., intake) of a nutrient and the physiological relevant outcome. The latter may be a biomarker of function, disease, or other health outcome (see Figure 8a.4). These may differ for a specific nutrient from one life-stage group to another because the critical function or the risk of disease may be different. EURECCA conducted a series of systematic reviews (Hooper et al., 2009; Van 't Veer et al., 2013; and Dhonukshe-Rutten et al., 2013)examining the intake-response relationship for several micronutrients in relation to health/clinical outcomes or surrogate outcomes associated with deficiency in order to establish a causal relationship between nutrient intake and a deficiency disease. Such relationships can be complex and influenced by many confounders and effect modifiers (e.g., ethnicity, life-stage group, body mass index, acute illness, and genotype etc). As noted earlier, surrogate outcomes are often used in RCTs when the study duration is too short to show an effect on health or clinical outcomes.

In addition to the focus on the prevention of nutrient deficiency, intake-response modeling can also be used to determine a safe upper intake level (i.e., ULs) (Section 8a.2.5) and chronic disease endpoints (Section 8a.3).

The final stage is to establish the two core NRVs for the nutrient(s) of interest and for the defined sex and life-stage group using the approach selected based on the available evidence. The two core NRVs are the Average Requirement (AR) together with the associated distribution of the requirement, and the UL. Strenuous efforts should be made to establish the AR for each nutrient because it is the basis for the multiple reference values; details are given below.

8a.2.4 Deriving the Average Requirement (AR) for a nutrient

The requirements for a specific nutrient vary from individual to individual and thus form a distribution of requirements. For most nutrients except iron, this variation in requirements is assumed to follow a normal symmetrical distribution as shown in Figure 8a.5, valid for a defined level of nutrient adequacy and a specific sex and life-stage group. Consequently, a coefficient of variation (CV) (i.e., the standard deviation divided by the mean × 100%) can be used to estimate the variation. In cases where requirements are not normally distributed (e.g., iron requirements for menstruating adolescent girls and women of child bearing age), data are transformed to achieve normality.

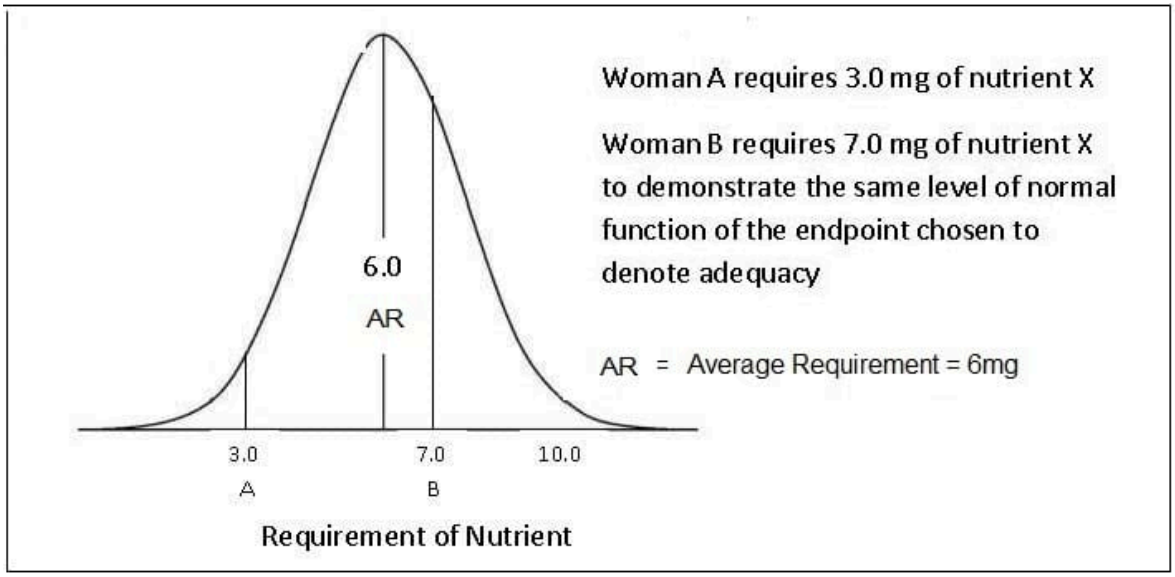


Figure 8a.5: Frequency distribution of the individual requirements (mg) of nutrient X in women 30–50y, reflecting variability in requirements among individuals. Adapted from King et al. (2007).

Inter-individual variability in requirements is affected by numerous diet- and host-related factors (Gibson, 2012). However, many of these factors are difficult to measure, or their impact is unknown. Hence, for many nutrients, the distribution of requirements is unknown, and instead is assumed to have a CV of about 10% (i.e., the standard deviation is about 10% of the mean requirement), assuming a normal distribution (King et al., 2007). EURRECA explored the biological variation in requirements associated with single nucleotide polymorphisms (SNPs) on the metabolism of five micronutrients (Claessens et al., 2013) but to date no information related to SNPs has been used in the derivation of NRVs due to a lack of relevant data.

The median of the requirement distribution represents the Average Requirement (AR) for that particular group of individuals (Figure 8a.6) and is used to assess the prevalence of nutrient adequacy within that group (see Chapter 8b for more details). Therefore, by definition the AR is:

> "the amount of nutrient that is estimated to meet the nutrient requirement of half the healthy individuals in a specific life-stage and sex group".

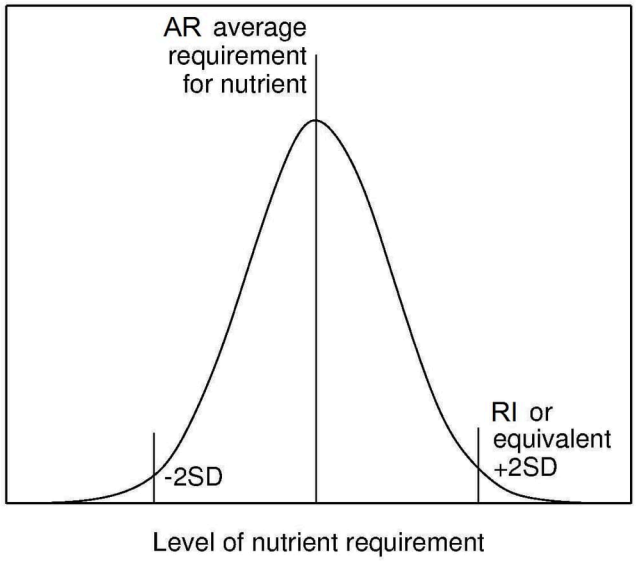


Figure 8a.6: Average requirement (AR) for a nutrient. The nutrient requirements are defined in relation to a frequency distribution of individual requirements. RI or the equivalent is defined as two standard deviations above the AR.

8a.2.5 Deriving the Safe Upper Levels of Intake

The Safe Upper Levels of Intake (ULs) represent daily intakes, that if consumed chronically over time, will have a very low risk of causing adverse effects. Intakes from all sources are considered: food, water, nutrient supplements, and pharmacological agents. The UL is defined as:

> "The highest level of habitual nutrient intake that is likely to pose no risk of adverse health effects in almost all individuals in the general population (King and Garza, 2007)."

The ULs are based on a toxicological risk assessment model involving a four-step process:

Box 8a.2: Four-step toxicological risk assessment model

- identification of risk of toxicity;
- dose-response assessment;
- assessment of the prevalence of intakes outside the reference values; and
- characterization of risks associated with excess intake

(WHO/FAO, 2005).

The dose-response assessment is built upon three toxicological terms: no-observed-adverse-effect level (NOAEL), lowest-observed-adverse-effect level (LOAEL), and uncertainty factor (UF). These terms are defined in Box 8a.3.

Box 8a.3: Dose-response assessment

- NOAEL is the highest continuing intake of a nutrient at which no adverse effects have been observed in the individuals or groups studied. In some cases, it may be derived from experimental studies in animals. When sufficient data are not available to define a NOAEL, then a LOAELis defined.
- LOAEL is the lowest continuing intake at which an adverse effect has been identified.

- UNCERTAINTY FACTORS (UFs) are used to address gaps in the data and incomplete knowledge (e.g., variability in response within the population; extrapolation from experimental animal to human data). The estimate for the magnitude of the uncertainty factor for each nutrient should be based on the approach of Renwick et al. (2004).

From (IOM, 1998).

At present, ULs are only set for those nutrients for which there is strong, high quality evidence. Like ARs, ULs vary by age and sex and tend to be lower for young children and pregnant women. The goal is to have less than 5% of a population subgroup with an intake greater than the UL, including intakes from supplements and fortified foods (IOM, 1998). This goal can sometimes be challenging for population subgroups consuming fortified cereals (Arsenault and Brown, 2003). For more details on the derivation of ULs, refer to WHO/FAO (2005).

This page titled 5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)

Before reviewing the NRVs established by each of the four individual authorities, it is stressed again that the process of setting nutrient reference values has evolved from setting a single value, equivalent to the RI, to specifying multiple reference values. In addition to establishing the two core NRVs — the Average Requirement (AR) and Safe Upper Level of Intake (UL) — for a specific nutrient and described in Sections 8a.2.4 and 8a.2.5, other NRVs are often specified and used for multiple purposes, some with a focus on optimizing health and preventing chronic diseases rather than the prevention of nutritional deficiencies. The more important of these other NRVs (the RI, LRI, and AI) are listed in Table 8a.1 and summarized below. The reader is advised to consult Table 8a.1 for a comparison of the recommended terminology with the terms currently used by these four authorities (Lupton et al., 2016).

In addition, the problems of setting NRVs for energy and some macronutrients, as well as NRVs for Chronic Disease Prevention, have also been faced by all four authorities. Comments on these difficulties are also considered below.

Nevertheless, many of the principles used previously to define NRVs remain the same (Box 8a.4); see Section 8a.2 for more details.

Box 8a.4: Underlying principles in defining NRVs

- Set for a particular group of individuals with specified characteristics, consuming a specified diet;
- Refer to the average daily need over a reasonable period of time, although the latter has seldom been defined; hence, the suggested amounts do not have to be consumed every day, but omissions or shortfalls must be balanced by increased intake on other occasions;
- Refer to levels of intake needed to maintain health in already healthy individuals; they do not allow for illnesses or stresses in life;
- Based on the typical dietary pattern of the country and may not be appropriate for persons following atypical diets;
- Consider possible interactions that involve nutrients and other dietary components and quantify them, if possible; and
- Assume that requirements for energy and all the other nutrients are met.

Recommended Intake (RI) (or equivalent) has been established by all authoritative groups. The RI is derived from the AR and its distribution (Figure 8a.6) and is used to assess intakes and plan diets for individuals. Currently, the convention has been to add 2SD to the observed AR to cover the needs of most (i.e., 98%) of individuals of the population. This means that an individual whose intake is equal to RI_{98} (i.e., AR + 2SD of the AR) has a 98% probability that their intake meets their needs (Figure 8a.6).

In the initial harmonization initiative in 2005 (King and Garza, 2007) it was suggested that in the future countries might wish to choose a lower percentile level deemed to represent an acceptable risk for nutrient inadequacy for an individual in their country instead of the conventional 98% (i.e., RI_{98}). To date, this suggestion has not been adopted.

Lower Reference Intake (LRI) (or equivalent) is also derived from the AR and its distribution and is set at 2 SD below the AR for each nutrient. This reference level represents the lowest intake that will meet the needs of some individuals. It has been established only by the UK and EFSA (Table 8a.1).

Adequate Intake (AI) (or equivalent) is applied when there is not enough information to establish an AR for a specific nutrient. The AI is defined as the observed or experimentally derived usual intake by a population group that appears to sustain health (King and Garza, 2007). The methods used to estimate the AIs vary. For example, for infants < 6mo, AIs are generally based on mean nutrient intakes (except vitamin D) supplied by human milk, whereas for children and adults the AIs can be derived from observed median intakes from national survey data, based on limited experimental data, or in some cases, epidemiological studies (Trumbo et al., 2013).

In general, mean usual nutrient intakes at the population level at or above the AI indicate a low probability of inadequacy. No assumptions can be made, however, about the prevalence of inadequate intakes when the mean intake of a group falls below the AI. Likewise, if an individual's usual intake equals or exceeds the AI, the diet is almost certainly adequate but again when the

usual intake of an individual falls below the AI, no estimate can be made of the probability of nutrient inadequacy (IOM, 2000). As noted previously, the Harmonization Committee urged that strenuous efforts should be made to establish an AR for each nutrient so that the inappropriate use of an AI can be avoided.

Energy Requirements are based on the average requirements (AR) for energy of individuals in a specified sex and life-stage group. The recommendations are not appropriate for the definition of requirements at the individual level. Adding an increment equivalent to 2SD to the average energy requirement would result in a recommendation that exceeds requirements and lead to overweight and obesity over the long term. Energy requirements are derived with the assumption that the requirements for all other nutrients are met (King et al., 2007).

The energy requirement is defined by FAO/WHO/UNU (2004) as:

> "The amount of food energy needed to balance energy expenditure in order to maintain body size, body composition, and a level of physical activity consistent with long-term good health. This includes the energy needed for optimal growth and development of children, for the deposition of tissues during pregnancy, and for the secretion of milk during lactation consistent with the good health of mother and child."

The criterion chosen on which to base the AR for energy is the total energy expenditure (TEE). When body weight and composition is stable in normal-weight individuals, the energy requirement is equal to TEE . In the past TEE for some life-stage groups was estimated indirectly by estimating the amount of energy consumed from self-reported food intakes and equating it with the amount of energy expended. This procedure was used by FAO/WHO/UNU (1985) and the United Kingdom (COMA,1991) to set the energy requirement for young children, but has now been abandoned in view of the concerns about the underestimation of energy intakes due to under-reporting; see Chapter 7 for more details.

The doubly labeled water method (DLW) method is the most accurate method for measuring TEE in free-living individuals and is now used by several authoritative groups including FAO/WHO/UNU (2004), the U.S. IOM (2005 and the UK SACN (2012); see Chapter 7 for details of the DLW method. TEE measured with DLW method includes the energy used to synthesize growing tissues but does not include the additional energy content of the tissue constituents (basically fat and protein) laid down during normal growth and pregnancy, or the milk produced during lactation.

Appropriate Macronutrient Distribution Range (or equivalent) serves as a guide to the distribution of percentage energy consumed as protein, fat, and carbohydrate said to be consistent with both the maintenance of health and a reduced risk of diet-related chronic disease, while providing adequate levels of essential nutrients. Increasingly, recommendations are being made for the contribution of free sugars as percentage of energy from total carbohydrate, as well as daily intakes of salt, dietary fiber, fruits and vegetables. (WHO, 2012; WHO, 2015), Although in general the range of intakes set are comparable, their meaning and application differ, with those of the United Kingdom (Section 8a.4.6), EFSA (Section 8a.6.6) and WHO (Section 8a.7.6) being population mean intake goals, whereas those established by the United States and Canada (Section 8a.5.6) are intended for individuals.

NRVs for Chronic Disease Prevention are being developed by several authoritative groups. However, several challenges have been encountered; these are itemized in Box 8a.5:

Box 8a.5: Challenges in defining NRVs for chronic disease prevention

- Relationships between diet and chronic disease are multifactorial in nature, and dependent on a variety of both nutritional and non-nutritional factors. These may include baseline risk for chronic disease by an individual, nutrient-diet or nutrient-nutrient interactions, and environmental and other lifestyle factors;
- Chronic diseases develop over a long period making it difficult to accurately measure dietary patterns over such a long period of exposure;
- Published intervention trials are often conducted in selected subgroups, such as individuals with an already increased risk of disease or history of disease, frequently with intervention dosages that are relatively high and which have relatively short follow-up periods.
- Evidence from observational studies can be compromised by uncertainty arising from problems of confounding, bias, and inaccuracy;
- Use of a validated dietary biomarker of nutrient exposure requires an estimate of the nutrient intake that corresponds with the clinical outcome;

- Challenges arise with the selection of meaningful and suitable outcome indicators for the chronic disease or an intermediate surrogate biomarker as a valid predictor of the chronic disease.
- Surrogate biomarkers are often required because the exposure and outcome are separated by such a long time period that clinical signs or symptoms may not be observed for many years. Criteria developed for a qualified surrogate biomarker are available in IOM (2010).

Adapted from Yaktine and Ross (2019).

A Chronic Disease Risk Reduction (CDRR) has been established for sodium in relation to the benfits of a reduced intake of sodium on cardiovascular disease risk, hypertension risk, systolic blood pressure and diastolic blood pressure. The CDRR for sodium was defined as:

> "The intake above which intake reduction is expected to reduce chronic disease risk within an apparently healthy population." (NASEM, 2019)

For further discussion of these challenges and proposed solutions, see Yetley et al. (2017).

This page titled 5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.4: U.K. Dietary Reference Values (8a.4)

The United Kingdom was the first country to adopt a frame-work for developing Nutrient Reference Values and these are termed "Dietary Reference Values" (DRVs) in COMA,1991; the frame-work used is depicted in Figure 8a.1. The generic term "Dietary Reference Values" was used to embrace three reference levels: the "Estimated Average Requirement" (EAR = AR), the "Reference Nutrient Intake" (RNI = RI), and the "Lower Reference Nutrient Intake" (LRNI, set at 2SDs below the AR). The term "reference values" was adopted in an effort to prevent users interpreting the figures as recommended or desirable intakes. Instead, the expert panel hoped that users would select the figure most appropriate for its intended use (Beaton, 1998).

UK Nutrient Reference Values for certain nutrients are not available for children < 5y of age. NRVs for thiamin and niacin equivalents were recalculated in 2011, based on the revised energy requirements established using a new approach (Section 8a.4.5) (SACN, 2012).

The NRVs for vitamin D for males and females aged 1–18y were also revised in 2016 (SACN) and now include a recommendation for selected life-stage groups to take a daily vitamin D supplement, especially from October to March.

A recent review of nutritional requirements of adults aged > 65y in the UK (Dorrington et al., 2020) has concluded there is evidence to support age-specific UK recommendations for those aged > 65y, for the RNIs for protein (1.2g per kg/d), calcium (1,000mg/d), folate (400µg/d), vitamin B_{12}(2.4µg/d) although to date, none of these suggested changes have been made by SACN (2016). In contrast, Dorrington et al. (2020) suggest that the current recommendations for the general population for sugars, dietary fiber and fatty acids, sodium and alcohol are probably appropriate for older adults.

8a.4.1 U.K. Estimated Average Requirement (EAR = AR) for nutrients

The term "Estimated Average Requirement" (EAR) in COMA,1991 represents the level of the nutrients estimated to meet the nutrient requirement of 50% of the healthy individuals in a particular sex and life- stage group. The EAR is especially useful for evaluating the possible adequacy of nutrient intakes of population groups.

8a.4.2 U.K. Reference Nutrient Intake (RNI)

Reference Nutrient Intakes (RNIs) were defined as 2SDs above the average requirement for each nutrient and represent the target for an individual's nutrient intake. The U.K. RNIs for minerals and vitamins are shown in Appendix 8a.1 and Appendix 8a.2 respectively. When data about variability in requirements were insufficient to calculate a SD, a coefficient of variation for the EAR of 10% was assumed. Habitual intakes above the RNI_{98} will be adequate for all but 2–3% of individuals in a specific sex and life-stage group. The RNI should not be used in relation to groups (COMA,1991).

8a.4.3 U.K. Safe Upper Levels for Nutrients

Safe upper levels were set by the U.K Expert Group on Vitamins and Minerals (EVM, 2003). They represents an intake that can be consumed daily over a lifetime without significant risk to health. Intakes from all sources were taken into account. The EVM group defined ULs for vitamin B_6, β-carotene, vitamin E, zinc, copper, selenium, boron and silicon, and also provided guidance for those nutrients for which the database was inadequate to establish a UL. Nutrients in this category included biotin, folic acid, niacin, riboflavin, pantothenic acid, thiamin, vitamin B_{12}, vitamin A, vitamin C, vitamin D, vitamin K, chromium, cobalt, iodine, manganese, molybdenum, nickel, tin, calcium, phosphorus, magnesium, iron, iodine, chromium, tin, and potassium. Suggested levels for these nutrients would not be expected to be associated with any adverse effects. Nevertheless, the EVM acknowledged that for those nutrients for which the database was inadequate to establish a UL, the suggested levels may not be applicable to all life stages or for lifelong intake, and should not be used as a UL.

8a.4.4 U.K. Additional Levels

Lower Reference Nutrient Intakes (LRNIs) were set by COMA,1991 at two standard deviations below the AR for each nutrient. The LRNIs represent the lowest intakes that will meet the needs of some individuals in the group. The LRNIs are used as a

monitoring tool for the UK National Diet and Nutrition Surveys. Habitual intakes below the LNRI are almost certainly inadequate for most individuals. For confirmation, however, biological parameters should be measured, especially when the nutrient intake of the individual lies between the LRNI and the EAR.

Safe Intake values were also set for some nutrients with important functions in humans, but for which the expert committee considered there were insufficient data to set DRVs (e.g., biotin, pantothenic acid, vitamin E, vitamin K, manganese, molybdenum, and chromium). Safe Intake was judged to be a level or range of intake at which there is no risk of deficiency and below a level where there is a risk of undesirable effects (COMA,1991).

8a.4.5. U.K. Average Requirement for Energy

In view of the increasing risk of overweight and obesity in the U.K., the SACN (2012) adopted a prescriptive approach and identified energy requirement values in relation to the best estimates of healthy body weights. Using this prescriptive approach, an overweight group consuming the amount of energy recommended for a healthy weight group are likely to lose weight whereas those underweight should gain weight (SACN, 2011).

The U.K average requirements for energy for adults of specified age, sex, and height assuming a median physical activity level (PAL) of 1.63 and expressed as MJ/d or kcal/d are available as tables in (SACN, 2011). Values given in the tables are derived from mean heights in 2009 for England (Health Survey for England 2009). These revised values apply to all adults unless energy expenditure is impaired, when a lower PAL value of 1.49 should be used. For children aged 1-18y, age and sex specific energy requirements, with PAL values ranging from 1.40 to 1.75, are presented, expressed as MJ/d or kcal/d. Separate tables for infants by age and sex who are breast-fed, fed breast-milk substitutes, mixed-feeding or unknown by age and sex are given, expressed as both MJ/kg or kcal/kg per day and MJ/d or kcal/day. The energy requirements for infants were adopted from FAO/WHO//UNU (2004).

The new U.K. factorial approach to establish energy requirements is based on the assumption that total energy expenditure (TEE) (or EAR) is equal to BMR × PAL (SACN, 2011). This approach was developed in recognition of the large unpredictable variation between individuals (inter-individual variation) in discretionary activity. This large inter-individual variation in discretionary activities is attributed in part to spontaneous physical activity (SPA). SPA includes body movements associated with activities of daily living, changes in posture, fidgeting, and a propensity for locomotion. An additional issue recognized by SACN (2011), is that PAL values, previously assumed to be independent of body weight, appear to have a complex relationship with body weight (Millward, 2012).

Conventionally, TEE is expressed as a multiple of Basal Metabolic Rate (BMR) and (PAL). Therefore:

$$\mathrm{TEE} = \mathrm{PAL} \times \mathrm{BMR}$$

Hence

$$\mathrm{PAL} = \mathrm{TEE} \ / \ \mathrm{BMR}$$

In this new approach, the PAL values were derived directly from DLW measurements of TEE in a reference population instead of applying PAL values predicted from lifestyle information, as used earlier by COMA in 1991 and 1994. Hence, to extract PAL values, TEE values measured by DLW in a suitable adult reference population were divided by BMR calculated from prediction equations of Henry (2005) and applying healthy body weights equivalent to a BMI of 22.5kg/m^2 and at the appropriate height of the adult population. from that same reference population (Tooze et al., 2007; Moshfegh et al., 2008). From the data on the distribution of PAL values in the adult reference population, PAL values for the 25th, median, and 75th percentiles corresponding to sedentary, low, and moderate activity were identified: i.e., 1.49; 1.63, and 1.78, respectively.

A separate 2006 data set for children was compiled from all DLW measurements of TEE of children aged over one year. The BMR for children was estimated using the equations of Henry (2005), applying healthy body weight based on the 50th percentile of the UK-WHO Growth Standards (aged 1–4y) and the 50th percentile of the UK 1990 reference for children and adolescents aged > 4y. From these data, PAL values for children aged 1–3y, 3 to less than 10y, and 10–18y were identified. To account for the cost of energy deposition during growth of the children (not included in TEE measured by DLW), the PAL values were adjusted by a simple +1% adjustment of PAL (PAL × 1.01). This results in acceptably low levels of error for children aged1–18y. Adjusted median PAL values for the three age groups were 1.40, 1.58, and 1.75. The reader is advised to consult SACN (2012) for more details.

8a.4.6 U.K. Adequate Macronutrient Distribution Range (AMDR)

The United Kingdom also proposed recommendations for carbohydrate, sugars, fats, and fatty acids. These are not defined in the same way as those for the micronutrients. They represent average intakes for populations and not for individuals, which are consistent with good health, and are expressed as a percentage of daily total energy intake and as a percentage of food energy (i.e., excluding the contribution from alcohol). Table 8a.3

Table 8a.3 UK dietary reference values (DRVs) and the WHO ranges of population nutrient intake goals (2003) for the energy supply
Figures are for percentage of total energy, unless otherwise stated. The U.S. and Canadian Acceptable Macronutrient Distribution Ra
[1] Recommendations apply to adults and children > 5y, unless otherwise stated.
[2] Represent the population average intake that is judged to be consistent with the maintenance of health in a population.
[a] Total fat includes all saturated and unsaturated fat (mono-and polyunsaturated)
[b] Saturated fat – several studies have shown a high saturated fat intake to be linked with high blood cholesterol. Elevated blood chol with unsaturated fat in the diet reduces blood cholesterol and lowers the risk of heart disease and stroke
[c] This is calculated as total fat–(saturated fatty acids + unsaturated fatty acids + trans fatty acids)
[d] Total carbohydrate includes all starch, sugars and dietary fiber
[e] Free sugars are sugars added to foods and drinks by the manufacturer, cook, or consumer, plus sugars naturally present in honey,
[f] Total fiber is the combination of Dietary Fiber, the edible, nondigestible carbohydrate and lignin components as they exist naturally health benefits.
[g] Women should not regularly drink more than 2-3 units of alcohol/day; Men – should not regularly drink more than 3-4 units of alco
[h] Target salt intakes set for adults (and children) do not represent ideal or optimum consumption levels, but achievable population g
** Based on SACN 2015 recommendations for the population ≥ 2y*
*** Based on COMA 1991 recommendations for the population ≥ 5y*
+ Based on WHO Guideline, 2015: Sugars intake for adults and children. Geneva, World Health Organization
++ Based on WHO.Guideline, 2012: Sodium Intake for Adults and Children; World Health Organization: Geneva, Switzerland.
From USDA U.S. Department of Health and Human Services Dietary Guidelines for Americans (2010) www.dietaryguidelines.gov
Based on Sodium Intake in Populations: Assessment of Evidence (2013) IOM of the National Academies Press

Dietary component	UK[1, 2]	
Total Fat*	Reduce to less than 35% of food energy[a] (excluding alcohol)	1
Saturated Fat	Reduce to less than 11% of food energy[b] (excluding alcohol)	<
PUFAs	6.5%	6
n-6 PUFAs (linoleic acid)		5
n-3 PUFAs α-linolenic acid		1
Trans fatty acids	< 2%	≤
MUFAs		B
Total[d] carbo-hydrate**	50% of food energy[d] with free sugars less than 5%[e]	5 fr p (2
Protein	15%	1(
Cholesterol	No specific recommendations	<

Dietary component	UK[1, 2]	
Dietary fiber	Adults 30g/d	A
Fruit and Vegetables	Increase to ≥ 5 portions (400g) of a variety of fruit and vegetables per day	≥
Alcohol	Should not provide more than 5% of energy in diet[g]	
Salt (Adults)	Not > 6g/d (2.4g Na)	N (2

summarizes the current U.K. recommendations for the macronutrients (COMA,1991; 1994; SACN, 2016) and compares them with the WHO (WHO/FAO, 2002) population nutrient intake goals for the prevention of diet-related chronic diseases (Section 8a.7.6). The U.S. and Canadian Acceptable Macronutrient Distribution Ranges (AMDRs) for adults (IOM, 2002) (Section 8a.5.6) are also given for comparison. Note that these U.S. and Canadian recommendations, unlike those set by the United Kingdom, EFSA, and WHO, are intended for use by individuals, as noted earlier.

In the UK, SACN (2016) included a recommendation for free sugars, stating that the average population intake of free sugars should not exceed 5% of total dietary energy for persons from 2y of age. They defined free sugars as those sugars added to food or those naturally present in honey, syrups and unsweetened fruit juices, but excluded lactose in milk and milk products. This report also stated that consumption of sugar-sweetened beverages should be minimized in children and adults in view of the evidence that consumption of sugar-sweetened beverages as compared with non-calorically sweetened beverages leads to greater weight gain and increases in BMI. SACN (2016) also adopted a broader definition of dietary fiber than that adopted earlier, and provided for the first time average population dietary recommendations for fiber intake for children and adolescents ranging from 15g/d(2–5y) to 30g/d (16–18y); see Pyne and Macdonald (2016) for more details.

This page titled 5.4: U.K. Dietary Reference Values (8a.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.5: United States and Canada- Dietary Reference Intakes (8a.5)

The United States and Canada have also adopted a paradigm for nutrient reference values that incorporates data to optimize health, prevent risk of chronic disease, and avoid deficiency. Their approach also provides multiple reference intakes for each nutrient to meet an expanding list of uses. Specifically, the U.S. Food and Nutrition Board has expanded their definition of requirements to:

- Consider reduction in the risk of chronic degenerative disease, where possible
- Establish an upper level of intake designed to avoid risk of adverse health effects, where data exist
- Include components of food not conventionally considered to be essential nutrients (eg choline, carotenoids, lycopenes, boron, vanadium) but which may have a possible benefit to health.

The nutrient reference values were established by a joint committee of both Canadian and U.S. scientists set up by the U.S. Food and Nutrition Board. The Committee adopted the generic term “Dietary Reference Intakes” (DRIs) for a set of reference values for each life-stage and sex group.

The criterion chosen by the U.S. Food and Nutrition Board to define nutrient adequacy differs for each nutrient and, sometimes, within the life-stage groups for the same nutrient; details are given in the separate reports published by the Food and Nutrition Board (1997, 1998, 2000, 2001, 2002) for 35 vitamins and minerals. Updated DRIs for calcium and vitamin D are given in IOM (2011).

8a.5.1 U.S. and Canadian Estimated Average Requirement (EAR = AR) for nutrients

This is defined as the median requirement for a specified criterion of adequacy for individuals of a certain life-stage and gender group. The specific criterion selected is defined by a specific function or biochemical measurement for each nutrient. Values for the EARs for nutrients are available in separate reports of the IOM (1997, 1998, 2000, 2001, 2002). When the mean usual intake of the group is equal to the EAR, 50% of the healthy individuals in that particular life-stage and sex group, meet their requirement and the other half of the group do not. Hence, usual intake at this level is associated with a 50% risk of inadequacy (Barr et al.,2003).

As noted earlier (Section 8a.2.4), each EAR refers to the average daily nutrient intake of apparently healthy persons over time, and this quantity does not have to be consumed every day. The EAR is especially useful for evaluating the possible adequacy of nutrient intakes of population groups. U.S. and Canadian EARs for elements and vitamins are included in Appendix 8a.3 and Appendix 8a.4 respectively.

8a.5.2 U.S. and Canadian Recommended Daily Allowance (RDA = RI)

This refers to the intake level that meets the daily nutrient requirements of almost all (≈ 98%) of the individuals in a specific life-stage and sex group. Values are given in the separate reports of the IOM (1997, 1998, 2000, 2001, 2002) (Barr et al., 2003). If the variation in requirement is well defined and symmetrically distributed, then the RDA is set at two standard deviations above the EAR: intakes at this level have a probability of adequacy ≈ 98%

$$\mathrm{RDA}_{98} = \mathrm{EAR} + (2\mathrm{SD})$$

If a coefficient of variation (CV) for the EAR of 10% is assumed, and as CV=SD/EAR, then.

$$\mathrm{RDA}_{98} = \mathrm{EAR} + (0.1 \times \mathrm{EAR}) + (0.1 \times \mathrm{EAR})$$

Or:

$$\mathrm{RDA}_{98} = 1.2 \times \mathrm{EAR}$$

Alternatively, if the CV is 15%, then:

$$\mathrm{RDA}_{98} = 1.3 \times \mathrm{EAR}$$

When the DRIs were first derived, a CV of 12.5% for protein, 15% for copper, molybdenum, and niacin, and 20% for both vitamin A and iodine was applied (King et al., 2007).

No RDA was proposed when there was not enough data to establish an EAR for that nutrient. Because the usual intake at the level of the RDA_{98} is, by definition, associated with a very low (2–3%) risk of inadequacy to an individual, the RDA_{98} is used as a recommended intake when assessing intakes and planning diets for individuals. For example, the appropriate target for phosphorus intake for a woman aged 31–50y is the RDA of 700mg. The RDA should not be used to assess the intakes of groups. RDA values for elements and vitamins are shown in Appendix 8a.3 and Appendix 8a.4 respectively.

8a.5.3 U.S. and Canada: Tolerable Upper Intake Level (UL) for nutrients

The Tolerable Upper Intake level (UL) for nutrients is the highest usual daily nutrient intake level likely to pose no risk of adverse health effects for almost all individuals in a life-stage and sex group. ULs have not been set for some nutrients with limited scientific data. Details of the adverse health effects used to set the ULs are given in the IOM reports (2000, 2001, IOM, 2011).

For some nutrients such as vitamin C, vitamin A, vitamin D, calcium, phosphorus, magnesium, copper, zinc, iron, selenium, iodine, and manganese, the UL refers to the total intake from all sources, including food, fortified food, water, supplements, and medications, where relevant. For others such as niacin and folate, the UL applies to forms from supplements, fortificants, and medications. The UL for magnesium represents intake from pharmacological agents only and does not include intake from food and water. In some cases, the form of the nutrient for the UL differs from that used for the RDA; examples include vitamin E, niacin, and folate.

The UL should be used by health professionals to ensure that nutrient intakes are not too high. As intake increases above the UL, the risk of adverse health effects increases. The UL is based on risk-assessment methodologies similar to those used in toxicological studies (1998). Figure 8a.1 shows the relationship of the observed level of intake to the risk of inadequacy and toxicity for the EAR, RDA, and the UL reference values.

8a.5.4. U.S. and Canada Additional Levels

Adequate Intake (AI) was also defined by the IOM and refers to a recommended average daily nutrient intake level based on observed or experimentally derived approximations or estimates of the nutrient intake by a group (or groups) of apparently healthy people. As noted earlier, for infants aged 0–6mo, the AI represents the nutrient intake values (except vitamin D) supplied by human milk (Allen et. al., 2018). The observed or derived intakes are assumed to be adequate. The AI is used when there are not enough scientific data to establish an AR (e.g., for fluoride, chromium, manganese, total fiber, and vitamins pantothenate, choline, and biotin), and is used as an intake goal for individuals. If the an individual's usual intake equals or exceeds the AI, their intake is almost certainly adequate, but if it falls below the AI, no estimate of the probability of nutrient adequacy can be made (IOM, 2000). Recently, AIs have also been set for sodium and potassium (NASEM, 2019).

8a.5.5 U.S. and Canada: Estimated Energy Requirements

Estimated energy requirements (EER) were compiled by the IOM, and are detailed in IOM (2005). The EER is defined as the average energy intake required to maintain current body weight and physical activity level (PAL) (and to allow for growth or milk production, where relevant) in healthy, normal-weight individuals of a specified age, gender, height, weight, and PAL. Tables provide values as well as equations based on height, weight, sex, age, and level of activity to predict TEE.

The EER values were based on DLW measurements of energy expenditure, and where relevant, the energy content of the tissue constituents (basically fat and protein) laid down in growing infants and children. When body weight and composition is stable in normal-weight individuals, the energy requirement is equal to total energy expenditure.

Several regression equations were developed for estimating the energy requirements of different life-stage and gender groups. Separate equations for individuals with a normal body weight (ie BMI 18.5–24.9kg/m^2) and those overweight or obese (BMI > 25kg/m^2) were developed. An example for the equation for normal-weight men aged 19y and above is given below:

$$\text{EER} = 661.8 - (9.53 \times \text{Age [years]}) + \text{PA} \times (15.91 \times \text{Weight [kg]} + 539.6 \times \text{Height [m]})$$

Where PA = physical activity coefficient corresponding to a given physical activity level (PAL). Table 8a.4. shows some sample calculations of EER for men with two different BMI values at three different heights and four different physical activity levels (PALs). Sample values for women are also available(Trumbo et al., 2002). More details of the four different physical activity categories are given in Table 8a.5.

Table 8a.4: Estimated energy requirements (EER) for men of 30y of age[1].

[1] For each year below 30, add 10kcal/d. For each year above 30 subtract 10kcal/d

[2] PAL = Physical activity level

[3] Derived from this regression equation based on doubly labeled water data: EER = 661.8−9.53×Age(y) ×PA×(15.91×Wt(kg)+539.6×Ht(m))

Where:

PA refers to coefficient for Physical Activity Levels (PAL).

PAL = total energy expenditure + basal energy expenditure

PA = 1.0 if PAL ≥ 1.0 < 1.4 (sedentary)

PA = 1.12 if PAL ≥ 1.4 < 1.6 (low active)

PA = 1.27 if PAL ≥ 1.6 < 1.9 (active)

PA = 1.45 if PAL ≥ 1.9 < 2.5 (very active)

Modified from Trumbo et al. (2002).

Height (m)	Weight (kg) for BMI of 18.5 kg/m^2	Weight (kg) for BMI of 24.99 kg/m^2	PAL[2]	EER[3] (kcal/d) BMI of 18.5 kg/m^2	EER[3] (kcal/d) BMI of 24.99 kg/m^2
1.50	41.6	56.2	Sedentary Low active Active Very active	1,848 2,009 2,215 2,554	2,080 2,267 2,506 2.898
1.65	50.4	1.50	Sedentary Low active Active Very active	2,068 2,254 2,490 2,880	2,349 2,566 2,842 3,296
1.80	59.9	81.0	Sedentary Low active Active Very active	2,301 2,513 2,782 3,225	2,635 2,884 3,200 3,370

*Table 8a.5: Description of physical activity level (PAL) categories. *Walking distance is estimated for the midpoint of the range (i.e., PAL = 1.5 for low active, 1.75 for active, and 2.2 for very active) and for midweight individuals weighing 70kg. Values in parentheses reflect walking distances at the same PAL for relatively heavy weight (120kg) and relatively light weight (44kg) individuals, respectively. Modified from Barr et al., Nutrition Reviews 61: 352–360, 2003.*

Physical activity category	PAL (multiples of basal energy expenditure)	Description	Physical Activity Coefficient (PA) M ≥ 19y F ≥ 19y
Sedentary	1.0 to < 1.4	Activities of daily living (ADL) only	1.00 1.00
Low active	≥ 1.4 to < 1.6	ADL plus walking about 2 miles/d (1.5/2.9)* or equivalent	1.11 1.12
Active	≥ 1.6 to < 1.9	ADL plus walking about 7 miles/d (5.3/9.9)* or equivalent	1.25 1.27

Physical activity category	PAL (multiples of basal energy expenditure)	Description	Physical Activity Coefficient (PA) M ≥ 19y F ≥ 19y
Very active	≥ 1.9 to < 2.5	ADL plus walking about 17 miles/d (12.3/22.5)* or equivalent	1.48 1.45

The physical activity coefficients (PAs) for these PAL categories vary slightly among the different regression equations, although the PAL for sedentary individuals is always 1.0 (Barr et al., 2003).

Note for individuals with a BMI ≥ 25, the estimated energy intake required to maintain current weight and activity level is termed the Total Energy Expenditure (TEE) and not the EER. This practice has been adopted because overweight is not consistent with long-term good health (Barr et al., 2003).

8a.5.6. U.S. and Canada Acceptable Macronutrient Distribution Range

The Acceptable Macronutrient Distribution Range (AMDR) of intakes is defined as a range of intakes for a particular energy source associated with reduced risk of chronic disease, while providing adequate levels of essential nutrients (IOM, 2002). They are intended for use by individuals, and have been established for carbohydrate, protein, total fat, n-6 poly-unsaturated fatty acids, and α-linolenic acid, as shown in Table 8a.3. Individuals should have intakes that fall within the limits of the AMDRs. If the usual intake of an individual is below or above the AMDR, there is a potential for increased risk of both inadequate intakes of essential nutrients and chronic diseases (IOM, 2002). Note that IOM have also defined an EAR and RDA for protein.

This page titled 5.5: United States and Canada- Dietary Reference Intakes (8a.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.6: European Dietary Reference Values (8a.6)

In 1993, the Scientific Committee for Food of the European Community (SCEC, 1993) proposed three reference values for each nutrient (Box 8a.6). subsequently, over a seven-year period, the European Food Safety Authority (EFSA) has generated separate publications on the NRVs for water, fats, carbohydrates, dietary fiber, protein, energy and 14 vitamins and 13 minerals, as well as a summary report (2017).

Box 8a.6 SCEC Dietary Reference Values

- Average Requirement
- Population Reference Intake
- Lower Threshold Intake

8a.6.1 European Average Requirements (ARs) for nutrients

The Average Requirement defined by EFSA is the mean requirement of a specific gender and life-stage group. The group was assumed to have a normal, symmetrical distribution, with the exception of the iron needs of menstruating women (2017). The EFSA ARs for vitamins, minerals, and trace elements for adults only shown in Table 8a.6. For the ARs for all life-stage groups for minerals and vitamins, see Appendix 8a.5 and Appendix 8a.6 respectively.

8a.6.2 European Population Reference Intake (PRI = RI)

EFSA (2017) defined "Population Reference Intakes" as the intake level adequate for virtually all people in a population group. It corresponds conceptually with the RI_{98}, and hence represents the AR + 2SD. The EFSA PRIs for vitamins, minerals, and trace elements for adults only are given in Table 8a.6 For the PRIs for all life-stage groups for minerals and vitamins, see Appendix 8a.7 and Appendix 8a.8 respectively.

For some life-stage groups where no data were available, interpolation or extrapolation was used by EFSA to set ARs and PRIs. The reference heights and body weights used by EFSA for children aged 0–2y are from the WHO Child Growth Standard (WHO, 2006), whereas for older children 2–18ythey use data from European children (van Buuren et al., 2012). Details of the reference body weights for children and adults used for scaling are given in (EFSA, 2017). Details of the criteria used to set the DRVs for infants (7–11mos, children (1–17y), adults, and pregnant and lactating women together with the type of studies used are also presented.

8a.6.3 European Tolerable Upper Intake Level for nutrients

The Tolerable Upper Intake Level (UL) was defined by EFSA as the maximum level of total chronic daily intake of a nutrient (from all sources — food, supplements, fortificants, water) judged to be unlikely to pose a risk of adverse health effects in humans (2017). ULs for six vitamins and eight minerals were established by SCEC in 2006 which are available in a summary report (EFSA, 2018). This report also provides updated information for the UL for vitamin D for infants aged 0–6mo and 6–12mo.

8a.6.4. European Additional Levels

*Table 8a.6 European Food Safety Authority Dietary Reference Values for adults. From EFSA (2017). Data for women, where different f megajoules; DFE: dietary folate equivalents. For combined intake of food folate and folic acid, DFEs can be computed as follows: µg f equivalent (1mg niacin = 1 niacin equivalent = 60mg dietary tryptophan); RE: retinol equivalent, 1µg RE = 1µg of retinol, 6µg of β-car *18–24y;** > 25y; *** AI (no AR has been set); ^ premenopausal; ^^ postmenopausal; Zn AR and PRI associated with 300, 600, 900*

Nutrient	Average Requirement (AR)	

Nutrient	Average Requirement (AR)	
Protein (g)	0.66 g/kg/BW/d	0.83 g
Vitamins		
Vitamin A (µg/d)	570 (490)	750 (6
Thiamin (mg/MJ)	0.072	0.1
Riboflavin (mg/d)	1.3	1.6
Niacin (mg NE/MJ)	1.3	1.6
Vitamin B_6 (mg/d)	1.5 (1.3)	1.7 (1.6
Cobalamin(µg/d)	4.0***	4.0
Vitamin C (mg/d)	90 (80)	110 (95
Folate (µg DFE/d)	250	330
Minerals		
Calcium (mg/d)	860*, 750**	1000*,
Phosphorus (mg/d)	550***	
Potassium (mg/d)	3500***	
Trace elements		
Iron (mg/d)	6 (7^, 6^^)	11 (16^
Zinc (mg/d)	7.5;9.3;11.0;12.7 (6.2;7.6;8.9;10.2)	9.4;11.7 (7.5;9.3
Copper (mg/d)	1.6*** (1.5***)	
Selenium (ug/d)	70***	
Iodine (ug/d)	150***	

Lower Threshold Intake (LTI) is one of two additional levels defined by the EFSA (2017) panel. The LTI is the intake below which almost all individuals in the population will be unable to maintain metabolic integrity according to the criterion chosen for each nutrient. The LTI represents the mean – 2SD.

Adequate Intake (AI) represents a value estimated when a Population Reference Intake cannot be established because an average requirement cannot be determined. An AI is the average observed daily level of intake by a population group (or groups) of apparently healthy people that is assumed to be adequate.

The EFSA (2017) report presents AI values for infants 0–6mo based on the nutrient supply from human milk. An AI is also given for those nutrients for which the data were deemed inadequate to set an AR (i.e., fluoride, iodine, manganese, molybdenum, phosphorus, potassium, selenium, copper, magnesium, biotin, choline, cobalamin, pantothenic acid, vitamin D, vitamin K). For a detailed discussion of these reference values, refer to EFSA (2017).

8a.6.5. European Average Requirement for Energy

The average requirements (ARs) for energy expressed as MJ/d for population groups within Europe are presented for specified age groups, and by gender for selected physical activity levels (PALs), depending on age (Table 8a.7); they are of limited use for individuals. ESFA also chose total energy expenditure (TEE) as the criterion on which to base the AR for energy for both adults and children. However, they determined TEE factorially from estimates of resting energy expenditure (REE), plus the energy needed for

various levels of physical activity (PAL) associated with sustainable lifestyles in healthy individuals. The equation is shown below:

$$\mathrm{TEE} = \mathrm{PAL} \times \mathrm{REE}$$

where PAL is a given physical activity level, and REE is resting energy expenditure. Table 8a.7.

Table 8a.7: ARs for energy (MJ/d) for adults. Note: The ARs for energy (1MJ = 238.83kcal) are calculated by multiplying estimates of predicted from anthropometric measures (Henry, 2005) with PAL values.
PAL values of 1.4. 1.6, 1.8, and 2.0 reflect low-active (sedentary), moderately active, active, and very active lifestyles, respectively.
Data from EFSA : ARs for infants and children and adjustments for pregnancy and lactation are also presented in EFSA (2013).

Age	PAL=1.4 M F	PAL=1.6 M F	PAL=1.8 M F
18–29y	9.8 7.9	11.2 9.0	12.6 10.1
30–39y	9.5 7.6	10.8 8.7	12.2 9.8
40–49y	9.3 7.5	10.7 8.6	12.0 9.7
50–59y	9.2 7.5	10.5 8.5	11.9 9.6
60–69y	8.4 6.8	9.6 7.8	10.9 8.8
70–79y	8.3 6.8	9.5 7.7	10.7 8.7

ESFA (2017) selected REE as a proxy for the slightly lower basal metabolic rate (BMR) because most studies estimated or measured REE. REE was estimated using the predictive equations of Henry (2005) which are derived from large datasets covering wide age groups. For adults 18–79y, the calculated REE was based on individual body heights measured in representative national surveys in 13 EU Member States, and body masses calculated from the measured heights assuming a BMI of 22kg/m^2.

PAL values were estimated from time-allocated lists of daily activities expressed as physical activity ratios available from the Compendium of Physical Activities compiled by Ainsworth et al.(2011) Examples of lifestyles associated with certain PALs estimated over 24h are tabulated in ESFA (2013).

To derive TEE as REE × PAL, PAL values of equal steps within the observed range of physical activity levels associated with a sustainable lifestyle were chosen for calculating AR for energy. In this way, PAL values could be allocated to lifestyles where values of 1.4, 1.6, 1.8 and 2.0, approximately reflect: low active (sedentary), moderately active, active, and very active lifestyles, were used, as shown in Table 8a.7.

No ARs for energy were calculated for adults > 80y in view of the paucity of European anthropometric data for this age group.

For infants from 0–6mos, energy requirements (i.e., ARs) were considered to be equal to the energy supply from human milk, and no AR was given. For infants aged 7–11mos, ARs were derived from TEE estimated by regression equations using DLW data, adding energy deposited in tissues for growth from the literature. Body masses for this age group were based on the WHO Child Growth Standards (WHO, 2006).

For children 1–18y, ARs for energy were based on predicted REE using the equations of Henry (2005) and using body masses and height from the WHO Growth Standards for children up to 2y, or from harmonized European growth curves for children 3–17y (van Buuren et al., 2012). PAL values of 1.4, 1.6, 1.8, and 2.0 were also applied for children. Energy expenditure for growth was accounted for by a 1% increase in PAL values for each of the three age groups (1–3y, > 3 – < 10y, and 10–18y).

For pregnant women, additional energy requirements were based on a mean gestational increase in body mass of 12kg, whereas for women exclusively breastfeeding during the first 6mos, the additional energy requirement during lactation was estimated factorially: see EFSA (2013) for more details.

8a.6.6 EFSA Reference Intake Ranges for Macronutrients

Reference Intake Ranges for Macronutrients have also been set by EFSA and are available in the 2017 summary report . The Reference Intake Ranges are defined as the intake range for macronutrients, expressed as a percentage of the energy intake and are intended as population averages. They apply to ranges of intakes that are adequate for maintaining health and associated

with a low risk of selected chronic diseases.

Table 8a.8 presents the macronutrients with a defined reference intake range expressed as the percentage of the energy intake for adults (> 18y) for total carbohydrate, total fat, linoleic acid (LA), and α-linolenic acid (ALA). A comparison with those set by the UK and WHO, also intended as population averages is presented in Table 8a.3. In contrast, the U.S and Canadian recommendations, also shown in Table 8a.3, are intended for use by individuals. Note for saturated fatty acids (SFAs) and trans-fatty acids (TFAs), an intake as low as possible is advised, whereas for dietary fiber and eicosapentanoic acid (EPA) plus docosahexanoic acid (DHA) , an AI is given. No reference intake ranges are given for protein, n-6 polyunsaturated fatty acids (PUFAs) or n-3 PUFAs, unlike WHO (see Section 8a.7.6). Likewise, no recommendation to restrict daily intake of cholesterol (mg/d) is stated, although an AI has been set for dietary fiber (g/d) which varies by age group. Note the data shown in Table 8a.8 can be used to assess the proportion of individuals in the population in a specified life-stage group with usual intakes outside the reference lower and upper limits of the range.

Table 8a.8: Reference Intake Ranges for total fat and carbohydrates and AIs for fatty acids and dietary fiber for adults ≥ 18y, and during pregnancy and lactation. From EFSA, 2017.

	Adults ≥18	Pregnancy	Lactation
Total Carbohydrates (E%)[a]	45–60		
Dietary fiber (g/d)[b]	25		
Total Fat (E%)[a]	20–35	20–35	20–35
SFA	ALAP	ALAP	ALAP
LA (E%)[b]	4	4	4
ALA (E%)[b]	0.5	0.5	0.5
EPA+DHA (mg/d)[b]	250	250	250
DHA (mg/d)[b]		+100–200[c]	+100–200[c]
TFA	ALAP	ALAP	ALAP

This page titled 5.6: European Dietary Reference Values (8a.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.7: WHO/FAO Nutrient Reference Values (8a.7)

When vitamin and mineral requirements for a specific country are not available, the most recent NRVs in the WHO/FAO (2004) report are often used. In this report WHO/FAO defined three levels shown in Box 8a.7:

Box 8a.7 WHO/FAO Nutrient Reference Values

- Estimated Average Requirement
- Recommended Nutrient Intake
- Tolerable Upper Nutrient Intake Level

8a.7.1 WHO/FAO Requirement (R = AR) for Nutrients

The requirement is defined by WHO/FAO as an average daily nutrient intake level that meets the needs of 50% of the "healthy" individuals of a particular age and gender. It is based on a given criteria of adequacy which varies depending on the specified nutrient. Where necessary, an allowance for variations in nutrient bioavailability has been included. However, a requirement has been set by WHO/FAO for only a few nutrients. In the interim, Allen et al.(2006) have calculated the ARs from the RNIs for a limited range of vitamins and minerals for infants, children and adults. The conversion factors used were based on SDs derived by the U.S. Food and Nutrition Board of the Institute of Medicine (FNB/IOM) for calculating the U.S and Canadian RDAs. They are tabulated in Annex C of Allen et al. (2006). These calculated ARs are shown in Appendix 8a.9.

8a.7.2 WHO/FAO Recommended Nutrient Intake (RNI = RI))

The Recommended Nutrient intake requirement is defined as a daily nutrient intake set at the requirement (AR) plus 2SD. This meets the nutrient requirements of almost all apparently healthy individuals in an age and sex-specific population group. Thus, it is equivalent to, and derived in the same way as individual requirements set by COMA (1991) for the U.K., by IOM (2001) for the U.S. and Canada, and by EFSA (2017) in Europe. When the standard deviation for the nutrient requirement is unknown, WHO/FAO (2004) has generally assumed a CV of 10%–12.5%, although there are exceptions for some nutrients (e.g., zinc, see below).

RNI values for dietary iron and zinc are based on estimates that meet the normative storage requirements, and are adapted from earlier reports (FAO/WHO, 1988; 1996). In these reports the normative requirement was defined as the mean requirement to maintain a level of tissue storage that is judged to be desirable. In Table 8a.9, the RNIs for zinc (mg/d) are given for diets with three levels of zinc bioavailability. For these derivations, a CV for the dietary zinc requirement of 25% is assumed. Appendix 8a.10 and Appendix 8a.11 summarize the mineral and vitamin Requirement Intakes (RNIs) of WHO/FAO (2004).

Table 8a.9: Recommended nutrient intakes for dietary zinc (mg/d) to meet the normative storage requirements from diets differing in zinc bioavailability. Unless otherwise specified, the within-subject variation of zinc requirements is assumed to be 25%.

[a] *For infants receiving maternal milk only, assumed CV = 12.5%. The bioavailability of zinc from human milk is assumed to be 80%.*

[b] *Formula-fed infants; moderate bioavailability for whey-adjusted milk formula and for partly breastfed infants or given low-phytate feeds supplemented with other milks; assumed CV = 12.5%.*

[c] *Formula-fed infants; low bioavailability applicable to phytate-rich vegetable-protein based formula with or without whole-grain cereal; assumed CV = 12.5%.*

[d] *Not applicable to infants consuming human milk only. From WHO/FAO (2004).*

Age group	Assumed weight (kg)	Bioavailability High	Mod.	Low
Infants & children				
0–6mo	6	1.1[a]	2.8[b]	6.6[c]
7–12mo	9	0.8[a]	—	—

Age group	Assumed weight (kg)	Bioavailability High	Mod.	Low
7–12mo	9	2.5[d]	4.1	8.4
1–3y	12	2.4	4.1	8.3
4–6y	17	2.9	4.8	9.6
7–9y	25	3.3	5.6	11.2
Adolescents				
F. 10–18y	47	4.3	7.2	14.4
M. 10–18y	49	5.1	8.6	17.1
Adults				
F. 19–65y	55	3.0	4.9	9.8
M. 19–65y	65	4.2	7.0	14.0
F. > 65y	55	3.0	4.9	9.8
M. >65y	65	4.2	7.0	14.0
Pregnant women				
1st trimester	—	3.4	5.5	11.0
2nd trimester	—	4.2	7.0	14.0
3rd trimester	—	6.0	10.0	20.0
Lactating women				
0–3 mo	—	5.8	9.5	19.0
3–6 mo	—	5.3	8.8	17.5
6–12 mo	—	4.3	7.2	14.4

8a.7.3 WHO/FAO Upper Tolerable Intake Level (UL) of Nutrients

The WHO/FAO Upper Tolerable Nutrient Intake Levels are defined for a limited range of nutrients as the maximum intake from food (including fortified products), water and supplements of that nutrient that is unlikely to pose risk of adverse health effects from excess in almost all (97.5%) apparently healthy individuals in an age and sex-specific population group. The ULs were developed using the model outlined in (WHO/FAO, 2006).

In October, 2017, the WHO held a technical consultation entitled: “Risk of excessive intake of vitamins and minerals delivered through public health interventions — current practices and case studies”. The proceedings of this technical consultation are available in a special issue. The reader is advised to consult the following for further details: (Pike and Zlotkin, 2019).

8a.7.4 WHO/FAO Additional Levels

WHO/FAO (2004) adopted additional terms for certain nutrients. For vitamin A, for example, they adopted the term “Recommended Safe Intake,” as was used in their earlier report (FAO / WHO, 1988),because of the lack of data for deriving a true mean requirement and variance for any group. The Recommended Safe Intake level is set to prevent clinical signs of deficiency and allow normal growth, but it does not allow for prolonged periods of infection or other stresses. As such, it represents the normative storage requirement plus 2SDs.

8a.7.5 WHO/FAO Human Energy Requirements

Energy requirements set by FAO/WHO (2004) are presented by gender and selected age groups and are expressed as energy (as kJ or kcal) per day and energy per kilogram of body weight. Requirements are also expressed as multiples of Basal Metabolic Rate (BMR). The recommendations for dietary energy intake are also accompanied by guidelines for desirable physical activity levels: see FAO/WHO(2004) for more details.

Note that, like the U.K and EFSA, the FAO/WHO energy requirements are specifically "prescriptive" in relation to body weights, defining suitable reference ranges consistent with long-term good health, amd should be applied to population groups and not to individuals.

The FAO/WHO (2004). energy requirements are based on estimates of TEE in free-living persons measured by DLW or other methods (e.g., individually calibrated heart rate monitoring) that give comparable results. They also modified their earlier factorial estimates of energy requirements which used occupational-related mean values of physical activity levels (PALs). Instead FAO/WHO (2004).used a range of life-style PAL values for each of the three lifestyles — sedentary, active, vigorously active, and shown in Table 8a.10.

Table 8a.10: FAO/WHO (2004) Classification of lifestyles in relation to the intensity of habitual physical activity, or PAL.
** PAL values > 2.40 are difficult to maintain over long periods*

Category	PAL value
Sedentary or light-activity lifestyle	1.40–1.69
Active or moderately active lifestyle	1.70 –1.99
Vigorous or vigorously active lifestyle	2.00–2.40*

The PAL values were calculated from the measurement of TEE and measurements or estimates of BMR (i.e., PAL = TEE/BMR). Both men and women are assigned to a PAL category using th same PAL values. The estimates of BMR were derived from the age and sex-specific predictive equations of Schofield (1985) using body weight alone.

The additional energy needs for pregnancy and lactation were also calculated using factorial estimates for the growth of maternal and fetal tissues, the metabolic changes associated with pregnancy, and the synthesis and secretion of milk during lactation.

For infants 1–12mos, the energy requirements were estimated from equations for TEE, adding the energy needs for growth. Separate energy requirements (as kJ/kg/d) are available for breastfed and formula fed infants with the same body weight up to 12mos of age because TEE is lower among breastfed than formula-fed infants during the first year of life.

For children and adolescents, measurements of TEE were derived using both DLW and heart rate monitoring, from which predictive quadratic polynomial regression equations for boys and girls were derived; equations are available in FAO/WHO (2004). The sum of TEE and the additional energy deposited in growing tissue and laid down during normal growth, represents the mean daily energy requirement (kcal or MJ/day). For more details, see FAO/WHO (2004).

8a.7.6 WHO/FAO Population Average Intake Goals for Macronutrients

The WHO ranges of population nutrient intake goals for preventing diet-related chronic diseases are shown in Table 8a.3. The recommendations are expressed as a proportion of the daily total energy intake rather than the absolute amount, with the exception of cholesterol and sodium, which are given in mg/d and g/d, respectively. This approach was adopted because the desirable up or down change will depend on the existing intakes in a given population (Nishida et al., 2004). The recommendations represent the population average intake that is judged to be consistent with the maintenance of health (i.e., low prevalence of diet-related diseases) in a population (IOM, 2002). Hence these focus on the maintenance of low *population risk* rather than low individual risk. In general, most of the ranges set are comparable to those of the U.K and the U.S. and Canada, although the meaning and application of the ranges differ. Exceptions are the recommendations by WHO/FAO for total fat intake which take into account countries with usual fat intake (as % energy) > 30% and < 15%. Total fat (as % energy) of at least 20% is consistent with good health.

The recommendation to restrict the intake of free sugars to less than 10% energy was made in view of the recognition that higher intakes of free sugars provide significant energy without supplying specific nutrients, and hence can have a negative impact on

diet quality. In addition, increasing or decreasing intake of free sugars is associated with parallel changes in body weight, a relationship that exists irrespective of the level of the intake of free sugars. The excess body weight associated with intake of free sugars arises from excess energy intake. Free sugars include monosaccharides and disaccharides added to foods and beverages by the manufacturer, cook, or consumer, and sugars naturally present in honey, syrups, fruit juices and fruit juice concentrations.

WHO emphasize that whole grain cereals, fruits and vegetables are the preferred sources of non-starch polysaccharides (NSP) (See Table 8a.3). They state that whole grain foods should together provide > 20g per day of NSP and > 25g/d of total dietary fiber. Note that fruits and vegetables do not include tubers (i.e., potatoes, cassava).

In addition to the population nutrient intake goals presented in Table 8a.3, WHO also emphasize the importance of maintaining at the population level an adult median BMI of 21–23kg/m^2, and for individuals a BMI in the range 18.5–24.9kg/m^2, with the avoidance of a weight gain greater than 5kg during adult life. Physical activity is also highlighted with a recommendation for a total of one hour per day on most days of the week of moderate- intensity activity , such as walking.

This page titled 5.7: WHO/FAO Nutrient Reference Values (8a.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

In light of the preceding discussion, it is not surprising that estimates for AR for nutrients and the RI_{98} derived from them vary among countries. Table 8a.11

compares the ARs for selected micronutrients set by the IOM in the U.S., by EFSA in Europe, and by COMA in the UK. Some of the discrepancies seen in Table 8a.11 arise because of differences in judgment by the expert groups setting the requirement estimates even when the same evidence-base has been consulted. Other potential sources of discrepancies in the AR and thus RI_{98} include:

- Disparities in philosophy about the most appropriate methodological approach to use
- Interpretation of data on which ARs are based
- Differences in selection of the criteria used to define nutrient adequacy
- Uncertainty in extent of metabolic adaption during pregnancy and lactation due to limited data
- Differences in number of life-stage groupings among countries
- Limited data on requirements for certain nutrients and life-stage groups
- Differences in scaling (both interpolation and extrapolation) for subgroups based on known data for other populations (e.g., adults)
- Varying bioavailability factors depending on the composition of habitual national diets
- Unknown factors influencing nutrient requirements.

Table 8a.11 A comparison of the ARs for male adults for selected micronutrients set by the US IOM in 1997–2011 (Otten et al, 2006; IOM, 2011), the European Food Safety Authority(EFSA, 2017), and the UK (COMA, 1991; SACN 2016).
**1 8–24y; ** > 25y*
[A] AI as no AR set. [B] Based on an EAR of 2550kcal.
NE niacin equivalents, DFE dietary folate equivalents.
For zinc, the EFSA sets three alternative ARs for levels of phytate intake above 300mg/d, of 600, 900, and 1200mg/d

Nutrient	IOM AR	EFSA AR	UK EAR
Vitamins			
Vitamin A (µg/d)	625	570	500
Vitamin D (µg/d)	10	15[A]	-
Vitamin C (mg/d)	75	90	25
Thiamin (mg/d)	1.0	0.072/MJ	0.80[B]
Riboflavin (mg/d)	1.1	1.3	1.0
Niacin (mg NE/d)	12	1.3/MJ	14.0[B]
Vitamin B_6 (mg/d)	1.4	1.5	1.2[B]
Vitamin B_{12} (µg/d)	2	4[A]	1.25
Folate (µg DFE/d)	320	250	150
Minerals			
Calcium (mg/d)	800	860*; 750**	525
Phosphorus (mg/d)	580	550[A]	400
Magnesium (mg/d)	350	350[A]	250

Nutrient	IOM AR	EFSA AR	UK EAR
Trace elements			
Iodine (µg/d)	95	150[A]	-
Iron (mg/d)	6	6	6.7
Selenium (µg/d)	45	70[A]	-
Zinc (mg/d)	9.4	7.5 9.3;11.0;12.7	7.3

Such discrepancies may be especially large for those nutrients and specific age groups for which the available data on requirements are very limited (e.g., children, adolescents, and the elderly). In such cases, requirements are often interpolated or extrapolated from data for other age groups, or they are not compiled at all. In addition, during pregnancy and lactation, maternal metabolic adaptation for certain nutrients may occur, but because the adaptation has not been firmly characterized, the additional nutrient needs are still equivocal.

A further source of discrepancy arises from the adjustments required that take into account the bioavailability of nutrients in the habitual national diet. The adjustments required depend on the nature of the diet ingested, the chemical form of the nutrient in the diet, and a variety of systemic factors known to affect the absorption and utilization of the nutrient. For many nutrients, factors affecting their bioavailability have yet to be established, so appropriate adjustments to yield dietary requirement estimates cannot be made. For others, fixed bioavailability factors are applied, even though the efficiency of absorption may vary with the dietary level of the nutrient or the life-stage group. For example, in U.K. diets, iron is assumed to have a fixed bioavailability of 15%, irrespective of the age and life-stage group (COMA,1991), whereas in the U.S. a factor of 18% is used except for women during the second and third trimester of pregnancy when a bioavailability factor of 25% is assumed (IOM, 2001). Other expert groups such as WHO/FAO (2004) employ differing factors to adjust for bioavailability, depending on the composition of the diet.

Several other factors besides sex, life-stage, and the habitual diet, are known to influence the requirements for many nutrients. Examples include body size, lean body mass, and activity level. For this reason, the requirement estimates are often set using a “standard” height and weight, and/or energy intake for a particular age and life-stage group; standards that may vary across countries. Therefore, those nutrients with requirements expressed per kg body weight or per MJ, may also differ. For example, reference heights and body weights used by EFSA (2017) are from the WHO Child Growth Standard (WHO, 2006) for children aged 0–2y, although for children 2–17y,data from European children are used (van Buuren et al., 2012). In the U.S., reference weights and heights for children and adults were based on anthropometric data collected from 1988–1994 as part of the Third National Health and Nutrition Examination Survey (NHANES III). However, the report on macro-nutrients (IOM, 2003) used data on median BMI and height-for-age from the CDC/NCHS growth charts (Kuczmarski et al., 2002).

In the future, factors such as race or ethnicity, lifestyle (e.g., vegetarians, smokers, oral contraceptive users), the existence of chronic disease (eg asthma, diabetes), environment, family history, and genetic predisposition to disease may also be taken into account when setting requirement estimates (Gibson, 2012). Examples of potential factors affecting requirements are summarized in (Figure 8a.7).

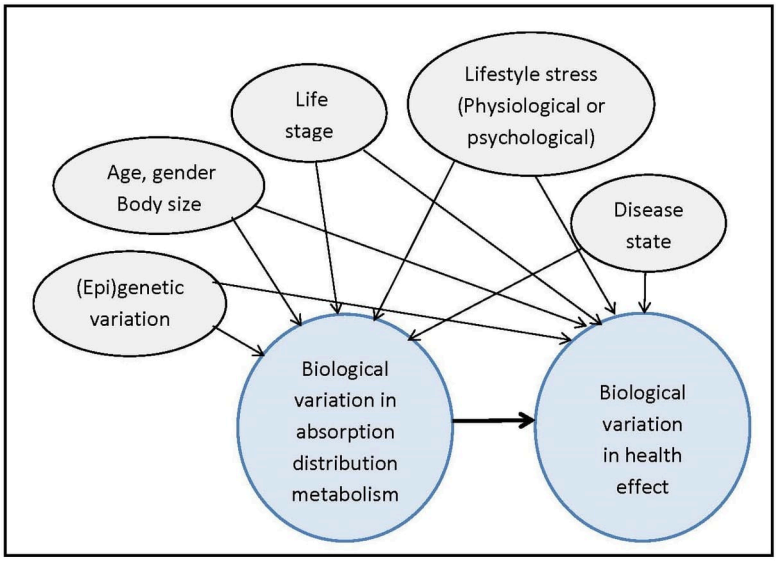


Figure 8a.7 Sources of biological variability of individual nutrient requirements. Modified from Ashwell et al., 2008.

In the United States and Canada, vegetarianism and some other lifestyle factors are already considered. As an example, the EAR for iron for vegetarians is higher than that for persons consuming a mixed Western diet to take into account the lower bioavailability of iron from a vegetarian diet (i.e., 10% vs 18%) (IOM, 2001), whereas the EAR for vitamin C for smokers is higher than for non-smokers (IOM, 2000).

The life-stage groupings are not defined in the same way among countries. North America has 22 such groups. Fewer groupings are defined in the U.K and Europe: Germany and the Netherlands have 14 each, but the number is presently not standardized, even within the European Economic Community(Trichopoulou and Vassilakou, 1990).

Finally, knowledge of the SD associated with the AR is required to set the RI_{98}. In many cases, however, the SD is calculated from the AR and an assumed CV, because the SD is unknown. Although a CV of 10% or 15% is often assumed, this is not always the case. WHO/FAO (2004), for example, has assumed a CV of 25% for the dietary zinc requirement estimate, resulting in a further source of discrepancy for the RI_{98} for zinc.

This page titled 5.8: Sources of discrepancies in Nutrient Reference Values (8a.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)

Nutrient Reference Values are used to assess the adequacy of nutrient intakes and for planning diets at both the individual and the popula tion level. They also have several important applications in food and nutrition policy, food regulation and trade. However, as highlighted above, the terminology and methodological approaches that are used for setting NRVs differ markedly among countries resulting in discrepancies in NRVs worldwide. Such disparate recommendations lead to international discrepancies in health, food policies and trade, highlighting important reasons for harmonizing the process for deriving NRVs globally; these reasons are summarized in Box 8a.8.

Box 8a.8 Reasons for harmonizing the process of deriving NRVs globally

- Improve objectivity and transparency of values derived by different national, regional, and international groups
- Provide a common basis or background for groups of experts to consider the processes leading to the development of the Nutrient Intake Values (NIVs)
- Allow for low-income countries with limited resources to convene groups of experts to identify how to modify existing reference values to meet the specific food supplies or international policies
- Provide a common basis across countries and regions for establishing global food and nutrition policies (i.e., fortification policies, regulatory issues)

From King and Garza (2007) and NASEM (2018b).

Recently, the feasibility of the new harmonized approach for deriving Nutrient Reference Values on a global scale has been confirmed following testing with three exemplar nutrients (zinc, iron, and folate) on two high-risk population subgroups — young children and women of reproductive age, as noted in Section 8.2. The success of the new framework depicted in Figure 8a.3, however, depends on six core values:

- NRVs are regularly updated
- The process is clear and transparent
- The methods are rigorous and relevant
- Factors influencing the NRVs are documented
- The strength of the evidence is determined
- The review is complete and efficient.

Several barriers have been identified that must be overcome before the proposed new organizing framework can be implemented globally. In addition, as highlighted earlier, many gaps in the scientific information required to compile NRVs still remain. This results in sources of uncertainty that must be considered, acknowledged and managed appropriately when deriving the NRVs.

This page titled 5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

6: EVALUATION OF INTAKES AND DIETS (CHAPTER 8B)

Abstract

This chapter addresses the assessment of nutrient intakes and the planning of nutritious diets at both the individual and population level. To assess intakes at the individual level for nutrients with an Average Requirement (AR) and Recommended Intake (RI), the likelihood of inadequacy can be assessed by using a statistical test that examines the individual's observed usual intake in combination with measures of variability of intakes and requirements. For nutrients with an Adequate Intake (AI) and Upper Intake Level (UL), statistical tests can be used to determine the level of confidence that an individual's observed usual intake is at or above the AI, or below the UL, respectively. Alternatively, published guidelines can provide a qualitative interpretation of the adequacy of an individual's usual intake of nutrients and macronutrients in relation to the Nutrient Reference Values (NRVs), and the Acceptable Macronutrient Distribution Range (expressed as a percentage of energy) (AMDR), respectively.

For assessment of intakes at the group level, the proportion with usual nutrient intakes below the AR and/or above the UL can be calculated by the probability approach or the simpler Estimated Average Requirement (EAR) cutpoint method. For nutrients with an AI that represents the mean or median intake of an apparently healthy group, when the group mean or median intake is at or above the AI, then it is likely that the prevalence of inadequate intakes for the group is low.

When planning nutritious diets for individuals, usual nutrient intakes should fall between the RI_{98} and the UL, or between the AI or UL. For energy and macronutrients, however, the targets for the diet are the individual's Estimated Energy Requirement (EER) and within the AMDR, respectively. For planning nutritious diets for groups, the current distribution of usual nutrient intakes should be altered to ensure a low prevalence of undesirable nutrient intakes, whereas for the appropriate energy level, the group mean energy intake must be close to the group mean EER. For the macronutrients, diets must be planned to reduce the prevalence of intakes outside the AMDR.

Dietary intake data generated from conventional dietary assessment methods have been severely criticized in the past because both random and systematic errors occur in measuring intakes of food and calculating their nutrient content (Subar et al., 2015). The data are also frequently misinterpreted. This is particularly unfortunate because the careful evaluation of food- and nutrient-intake data has important uses and applications, provided that the limitations of the methods are clearly understood.

This chapter describes the assessment of nutrient intakes and the planning of nutritious diets at both the individual and population level. Most of the approaches used are based on comparison of the usual daily intake of an individual or group with the appropriate Nutrient Reference Values (NRVs) or the Acceptable Macronutrient Distribution Ranges (AMDRs). The tables of NRVs have now been expanded in many countries and regions to include:

- An Average Requirement (AR), referred, in some countries, as an Estimated Average requirement (EAR).
- A Recommended Intake (RI_{98}), termed a Recommended Dietary Allowance (RDA) in Canada and the U.S.
- An Adequate Intake (AI) and/or an Upper Intake Level (UL),

Each NRV has a specific use as a reference standard for diet assessment and planning; their derivation is described in Chapter 8a. The AI is used for those nutrients for which there is insufficient available information to set an AR. Development of the UL arose because of the increasing availability of fortified foods and dietary supplements worldwide, and concern about ingesting too much of a nutrient.

6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

6.4: Using the NRVs for planning nutritious diets (8b.4)

This page titled 6: Evaluation of intakes and diets (Chapter 8b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

Many countries have adopted a new paradigm and set multiple NRVs rather than just a single value for the Recommended Intake (RI_{98}) or equivalent (Chapter 8a). Nutrient Reference Values are used for assessing the adequacy of nutrient intakes and planning nutritious diets for both apparently healthy individuals and population groups as shown in Figure 8b.1. However, care must be taken to ensure the appropriate NRV is selected and the data interpreted correctly.

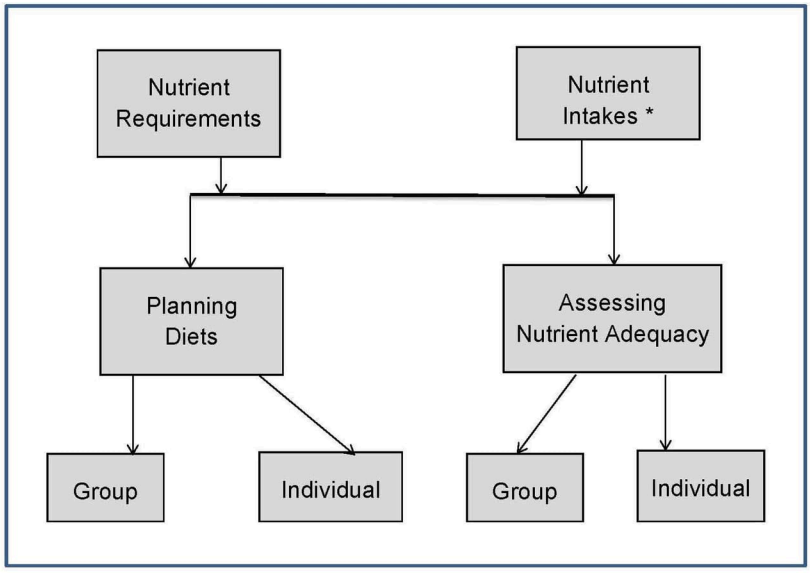


Figure 8b.1. Conceptual framework — uses of Nutrient Reference Values (NRVs). * Food plus supplements. Adapted from Beaton (1994).

Assessment of nutrient intakes and planning nutritious diets at the individual level is especially challenging because neither the requirements of the individual nor their usual nutrient intake are known with any certainty. A quantitative method has been developed to estimate the probability that an individual's intake or diet meets his or her requirement. This method can be used provided NRVs for the nutrient of interest exist and the day-to-day variation in intake of the nutrient is not high.

Alternatively, health professionals can use published guidelines to provide a more practical qualitative assessment based on the likelihood that the intake of an individual is not adequate. Both methods can be used for dietary counseling and nutrition education and for developing a dietary plan that the individual can follow.

At the group level, with the inclusion of both an Average Requirement (AR) and Upper Intake Level (UL) in the expanded NRV framework, the focus is no longer on the group mean intake alone, but also takes into account the distributions of usual intake within a group. With this new focus, the prevalence of inadequate and potentially excessive intakes can be assessed by estimating the proportion within the group whose usual intakes are less than the AR, and above the UL, respectively, by using either the probability approach or the simpler EAR cutpoint method. To ensure a nutritious diet for the group, a plan can be devised to alter the current distribution of usual intakes to reduce the prevalence of undesirable intakes.

Guidance on the correct uses of NRVs for planning and assessing nutrient intakes of individuals and groups to avoid their misuse is summarized in Table 8b.1. A more detailed discussion of the correct use of each component of the NRVs for the assessment of nutrient intakes and planning nutritious diets of individuals and groups follows.

Table 8b.1. Uses of Nutrient Reference Values for planning and assessing nutrient intakes of apparently healthy individuals and groups. Modified from Barr (2006).

NRV	Assessment of nutrient intakes of healthy individual or groups
AR Individual	Use (with information on variability of requirement and intake) to examine the probability that the usual intake is inadequate
AR Group	The proportion of the group with usual intake below the AR is an estimate of the group prevalence of inadequacy
RI Individual	Usual intake at or above the RI has a low probability of inadequacy
RI Group	Do not use to assess intakes of groups
AI Individual	Intake at or above the AI can be assumed adequate; no assessment can be made if intake is below the AI
AI Group	Median usual intake at or above the AI implies a low prevalence of inadequate intakes; no assessment can be made if median intake is below the AI
UL Individual	Usual intake above the UL may place an individual at risk of adverse effects from excessive intake
UL Group	The proportion of a group with usual intake above the UL may be at potential risk of adverse effects of excessive intake
NRV	**Planning for nutritious diets for healthy individuals or groups**
AR Individual	Do not use the AR as an intake goal; usual intake at this level has a 50% probability of inadequacy
AR Group	Plan for an acceptably low proportion of a group with intakes below the AR. Note that mean intake will likely be above the RI
RI Individual	Aim for this intake; usual intake at or above the RI has a low probability of inadequacy
RI Group	Do not use the RI to plan mean intakes for groups. In almost all cases, mean intake at the RI will lead to an unacceptably high prevalence of inadequate intakes
AI Individual	Aim for this intake; usual intake at or above the AI has a low probability of inadequacy
AI Group	Plan for median intake at this level; median usual intake at or above the AI implies a low prevalence of inadequate intakes
UL Individual	Plan for usual intake to remain below the UL to avoid potential risk of adverse effects from excessive intake
UL Group	Plan to minimize the proportion of a group with intakes above the UL to minimize the risk of adverse effects of excessive intake

With the development of more reliable statistical methods to assess the risk of inadequate or excessive intakes of nutrients, vulnerable individuals and subgroups in the population can be more accurately identified and targeted during intervention programs. Table 8b.2 summarizes some of the uses and health applications that depend on the NRVs using examples from the United States, Canada, and the United Kingdom.

Table 8b.2. Critical health applications that depend on the NRVs in the United States, Canada, and the United Kingdom.
WIC: U.S Special Supplemental Nutrition Program for women infants, and children.
SNAP: U.S Supplemental Nutrition Assistance Program.
Data from Murphy et al. (2016).

Application	Examples
Dietary guidelines	Important in formulating food-based dietary guidance, e.g.: • US Dietary Guidelines for Americans • Eating Well with Canada's Food Guide (2007) • USDA Food Patterns • UK Eatwell Guide
Nutrition monitoring	Needed to asses nutritional health on a national level. • US NHANES and What We Eat in America analyses • Canadian Community Health Survey analyses • UK Diet and Nutrition Survey Rolling Program
Food assistance programs	Important to guide the design of healthier federal nutrition assistance programs • School meals, WIC, SNAP, child and adult care programs • Administration on Aging programs
Health Professionals	Used for dietary counseling and education and to design: • healthy diets for institutions such as hospitals • healthy diets for long-term care facilities and prisons
Nutrition research	Needed to study how diet can help prevent disease and provide a frame of reference in research
Military	Used to • Ensure nutrient needs are met for armed forces • Plan healthy meals • Procure food, including military rations
Nutrition labeling	Can be used for the nutrition facts supplement facts labels. Such labels can help consumers make healthier food choices
Food/supplement industries	Used to develop healthy foods and safe supplements
Global nutrient standards	Provide a framework for other countries and international organizations when setting their own standards

This page titled 6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

None of the methods of assessing nutrient-intake data described below are capable of identifying with certainty that an individual's recorded nutrient intake is adequate or excessive. This is because the actual nutrient requirement of an individual is not known. In addition, the recorded nutrient-intake data only approximate the individual's "usual" nutrient intake, because of normal day-to-day variation in the diet (Chapter 6) combined with measurement errors (Chapter 5). For these reasons, dietary data alone can only provide an estimate of the probability of an inadequate or excessive intake of a nutrient. The reliability of this risk estimate depends on the quality of the dietary data collected and the method used to calculate the estimate of risk. However, only when biochemical, anthropometric, and clinical assessments are combined with the dietary investigation can a valid assessment of an individual's nutritional status be made.

Table 8b.1 summarizes the appropriate Nutrient Reference Values for assessing the nutrient intakes of individuals. Details on planning diets for individuals are given in Section 8b.4.1. The outline in this table should always be followed when assessing nutrient intakes (and planning diets) for individuals.

8b.2.1 Using the Average Requirement

It is not possible to compare the intake of an individual with his or her own requirement, because the actual requirement for any given individual is not known, as noted earlier. Moreover, the reported nutrient intake of an individual is unlikely to represent their mean usual intake because of within-person variation in nutrient intakes combined with measurement errors, as discussed in Chapters 5 and 6. In view of these uncertainties, two approaches have been developed to assess the nutrient intakes of an individual. The first approach involves a statistical method that estimates the *level of confidence* that the usual nutrient intake of an individual meets the corresponding AR and is discussed below. The second approach is a qualitative interpretation of the nutrient intakes of an individual and is described in Section 8b.2.4.

The statistical approach is based on the following assumptions:

- The AR is the best estimate of an individual's requirement.
- There is person-to-person variation in the requirements (i.e., between-person variation), best described by the standard deviation of the requirement estimate (SD_R). The latter indicates how much the individual's requirement for a nutrient can deviate from the median requirement in the population. In the absence of data for SD_R, IOM (2003) has assumed a CV of 10% for most nutrients. The SD_R is then calculated from the AR and CV as: $SD_R = CV \times AR$.
- Mean reported intake of an individual is the best estimate of the usual intake of an individual.
- Day-to-day variation in intake (i.e., within-person variation) for an individual exists and can be described by the within-person standard deviation of intakes (SD_{within}). The latter indicates how much the reported intake may deviate from the usual intake.

The statistical approach standardizes the difference between the reported intake of a young woman for zinc, as an example, and the AR for zinc, by dividing the difference by the standard deviation of the difference (SD_D) . The latter takes into account the day-to-day variability in zinc intake for the young woman under study, and the variability of the requirement (her zinc requirement could differ from the AR) (Barr et al., 2002; Murphy and Poos, 2002). For details of the equation and worked examples using the ARs, see Appendix B in IOM (2000). Note this statistical approach cannot be used when the within-person variation is not normally distributed, or when the requirement distribution is skewed (e.g., iron requirements for premenopausal women).

The result of this calculation is a Z-score. In general, Z-scores represent a difference divided by its standard deviation, and correspond to probability values associated with the normal distribution. Selected Z-values corresponding to different levels of assurance are given in Table 8b.3). From the Z-score the probability value (p) reflects the degree of confidence that, for example, the usual zinc intake of the young woman meets her zinc requirement. As shown in Table 8b.3) when Z = +0.50, p = 0.70; when Z = −1.0, then p = 0.15, and when Z = −2.0, then p = 0.02 (Barr et al., 2002).

Standardized Difference (Diff/SD_D) i.e. Z-score	Interpretation
+ 2.0	Usual intake is adequate
+ 1.65	
+ 1.5	
+ 1.0	
+ 0.5	
0.0	
– 0.5	
– 1.0	
– 1.5	
– 1.65	
– 2.0	

8b.2.2 Using the Adequate Intake

The approach described in Section 8b.2.1 cannot be used for those nutrients for which there is insufficient available information to set an AR. Instead, a statistically based hypothesis testing procedure can be used to compare an individual's intake to the AI. Details are also given in IOM (2000). Again, the procedure consists of a Z-test involving the difference between the intake of the individual and the AI, relative to the estimated within-person standard deviation of the daily intake of the nutrient of interest.

If there is a high degree of confidence that the usual nutrient intake of an individual equals or exceeds the AI, after applying the appropriate statistical test, the diet is almost certainly likely to be adequate in a given nutrient. In contrast, no evaluation can be made of the probability of an inadequate nutrient intake when the intake of the nutrient falls below the AI because the requirement is unknown. In such cases, the interpretation must rely solely on the judgment of nutrition professionals. In general, to eliminate the possibility of nutrient inadequacy, an individual should be encouraged to increase his or her nutrient intake to meet the AI (Barr et al. 2002; Murphy and Poos, 2002).

8b.2.3 Using the Upper Intake Level

The Upper Intake Level (UL) can be used to determine whether the usual intake of an individual is so high that it poses a risk of adverse health effects. Another statistical test can be used to determine the level of confidence that the usual intake of an individual is below the UL; details of this test are also given in (IOM, 2000). Whether the UL applies to the intake from supplements, fortificants, or medications or to the total intake from all sources depends on the nutrient (Barr et al. 2002; Murphy and Poos, 2002).

Hypothetical examples of assessing the adequacy or excessive intakes of riboflavin, folate, calcium, vitamin D and zinc for an elderly woman are given in Barr et al. (2002). In these examples, the appropriate statistical methods were employed to determine the level of confidence that the woman's usual intake met her requirements (for riboflavin, folate, and zinc based on an AR), or exceeded the AI (for calcium and vitamin D at that time), but was below the UL; see Barr et al. (2002) for more details. Note that an AR has now been set for calcium and vitamin D by IOM (2011).

8b.2.4 Using the qualitative interpretation of an individual's nutrient intake

Table 8b.4: Guidelines for the qualitative interpretation of individual intakes. Modified after IOM (2003).

Recommended Intake (RI_{98})
If the intake is > RI_{98} there is a high level of confidence that intake is adequate if observed over a large number of days

	Recommended Intake (RI_{98})
If the intake is between the Average Requirement (AR) and the RI_{98}, the intake probably needs to be improved because the probab	
If the intake < AR, then the intake very likely needs to be improved because the probability of adequacy is less than 50%	
Adequate Intake (AI)	
If the intake > AI, then the mean intake is likely adequate if observed over a large number of days	
If the intake < AI, then the adequacy of intake cannot be determined	
Upper Intake Level (UL)	
If the intake > UL, there is a potential risk of adverse effects if observed over a large number of days	
If the intake < UL, the intake is likely safe if observed over a large number of days	

For practical purposes, a qualitative interpretation of the nutrient intakes of an individual in relation to the AR, RI_{98}, AI, and UL can be made, using the guidelines given in Table 8b.4. To apply this approach, the intakes of an individual must be assessed over a large number of days. Again, it is preferable to combine such a qualitative assessment with information based on biochemical, clinical, and anthropometric data, where possible.

8b.2.5 Using the Acceptable Macronutrient Distribution Range

If Acceptable Macronutrient Distribution Ranges (AMDRs) have been set for individuals for carbohydrate, protein, total fat, linoleic acid, and α-linolenic acid, e.g., by IOM (2003) (see Section 8a.5.6), then the usual intake of macronutrients (expressed as a percentage of total energy intake) for an individual should fall within these ranges. If the usual intake of a macronutrient is below the AMDR, there is potential for an increased risk of both inadequate intake of the essential macronutrient and chronic diseases. The potential for an increased risk of chronic diseases may also occur if the usual intake of a macronutrient exceeds the AMDR (IOM, 2000).

8b.2.6 Using Energy Intakes

The average estimated energy requirements (EER) are based on a defined age, sex, height, weight, and physical activity level. Hence, a mean usual energy intake for an individual either above or below their calculated EER, would theoretically result in either weight gain or weight loss, respectively. However, even based on several days of energy intake from an individual, it is difficult to determine energy balance. Therefore, in practice recent body weight is often used to assess the likely adequacy of the energy intake of an individual(Barr, 2006). Note that no probability of adequacy can be assessed for energy, because the intake of energy is almost always related to energy requirements, violating one of the assumptions of the probability approach (see Section 8b.3.1).

This page titled 6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

Information is often required on the proportion of a group with usual intake of a nutrient below their own requirement. An estimate of the proportion of the group with an excessive usual nutrient intake is also important, as it exposes that portion of the group to the risk of adverse effects (See Table 8b.1). Unfortunately, there are some examples in the literature in which the use of the NRVs to determine the prevalence of inadequate intakes has been incorrect, leading to misleading interpretations (Trumbo et al., 2010). These may include:

Comparison of the usual intakes of the individuals with the RI_{98}. Reporting the proportion of a group with usual intakes below the RI_{98} is inappropriate because it will always lead to an overestimate of the true prevalence of inadequacy. By definition, the RI_{98} is set at 2SD above the AR, and thus exceeds the requirements of more than 97% of all individuals in the group.

Comparison of the group mean or median intake with the AR. This is inappropriate because even if the mean intake equals the AR, a high proportion of the population (≈ 50%) will be expected to have inadequate usual intakes based on the definition of the AR.

Comparison of the group mean or median intake with the RI_{98}. This is inappropriate because it does not take into account the wide variation in the distribution of usual intakes within the group, which almost always exceeds the variability in the requirement distribution. Hence, even if the mean/median intake equals or even exceeds the RI_{98}, there may still be a substantial proportion of the group who will have intakes less than their own requirements, as shown in Box 8b.1. Indeed, to ensure a low prevalence of intakes below the AR, the mean or median nutrient intakes of the group should exceed the RI_{98}, often by a considerable amount.

Comparison of the usual intakes of individuals with the AI . This cannot be used to assess the prevalence of inadequate intakes.

Comparison of the group mean or median intake with the AI. Provided the AI used has been set based on the mean intake of a healthy population, then if the mean or median intake of the group is at or above the AI, the group can be assumed to have a low prevalence of inadequate intakes. However, if the mean or median intake of the group is below the AI, it is not possible to make any assumptions about the prevalence of inadequacy.

Box 8b.1. Inappropriate use of the RI_{98} to assess group mean intakes

Percentile	5th	10th	15th	25th	50th	75th	85th	90th	95th
Vit. B-6 intake (mg/d)	0.92	1.02	1.11	1.24	1.51	1.90	2.13	2.31	2.65

- The data above show selected percentiles of the distribution of vitamin B_6 intakes for women 51–70y adjusted for within-person variability from U.S National Health and Nutrition Examination Survey III 1988–1994 (NHANES III). The US AR for vitamin B_6 is 1.3mg/d and the U.S RI_{98} is 1.5mg/d for women of this age group.
- When the median intake (1.51mg/d) for the group is compared with the RI_{98} of 1.5mg/d, it may seem that inadequate intake of vitamin B_6 is not a problem.
- However, when the distribution of usual intakes is compared with the AR cutpoint, the AR value (1.3mg/d) falls between the 25th percentile and the 50th percentile of usual intakes. Hence, more than 25 percent of usual intakes are below the AR cutpoint.
- The result is that the prevalence of inadequacy within the group is greater than 25 percent (but less than 50 percent).

Adapted from Otten et al. (2006).

Note that in cases when the mean or median intake is compared with the NRVs, the distribution of nutrient intakes does not need to be adjusted for day-to-day variation because only the mean or median intake is being examined. However, the only situation in which this approach is appropriate is for nutrients with an AI based on the mean/median intake of a healthy population, or for

energy (see Section 8b.4.2).

When the objective is to compare the adequacy of nutrient intakes for two groups, then the prevalence of inadequate intakes for each group should be determined, and not the mean intake for each group. The latter can be the same as shown in Figure 8b.2, even though there were differences in the proportions in each group with inadequate intakes. This situation occurs when the intakes in one group are much more variable than in the other group (Murphy and Poos, 2002, Figure 8b.2).

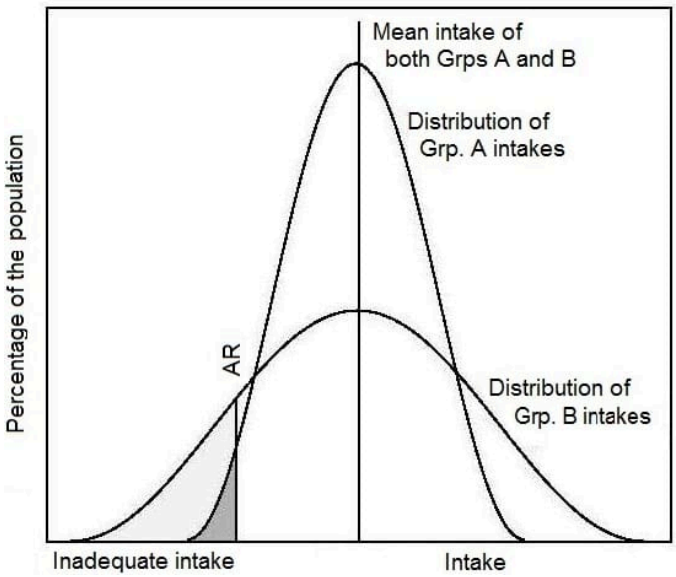


Figure 8b.2: Groups A and B have the same mean intake, but the intakes of Group A are less varied. As a result many fewer subjects in Group A have intakes below the AR (dark shaded area), relative to Group B (total shaded area on low-side of AR). Modified from Murphy and Poos (2002).

The recommended approach for evaluating the adequacy of nutrient intakes of population groups is to use the average requirement (AR). Two methods based on the AR have been developed: the full probability approach and the EAR cutpoint method; they are described below. Neither of these methods requires that the intakes be normally distributed, although other assumptions described below must be satisfied (Murphy and Vorster, 2007).

Note that these methods cannot be used to identify actual individuals with inadequate intakes because some of the individuals will have requirements that are lower than the AR, and others have higher than average requirements.

8b.3.1 Full Probability Approach

The full probability approach was first described by Beaton (1972), with the goal of providing a more reliable estimate for the proportion of individuals within a population group with inadequate intakes (i.e., intakes below their requirements). Because there is no information about the actual requirements of each individual, this procedure does not identify with certainty which individuals have inadequate intakes. Hence, it cannot be used to screen those individuals for nutrient inadequacy.

The statistical approach combines the distributions of both requirements and individual usual nutrient intakes for the group to estimate the proportion of individuals with inadequate intakes. The probability of inadequacy for each individual in the group is calculated first, followed by an estimate of the average of the individual probabilities for the group. This yields the group prevalence of inadequate intakes. The statistical approach is based on three key assumptions:

- Intakes and requirements are independent
- The median and the variance of the distribution of requirements are known
- The shape of the requirement distribution follows some known statistical model (normal, lognormal etc) (IOM, 2003).

To adopt the probability approach for evaluating nutrient intakes, the following information is required:

- The AR for each nutrient for the particular sex and life-stage group of individuals.
- The distribution of requirements for each nutrient among similar individuals: for most nutrients, this is not precisely known. In the absence of such information, the distribution of requirements for most nutrients is assumed to be normal with a coefficient of variation of 10% (or 15% for niacin) about the AR. Notable exceptions are the iron requirement distributions for menstruating adolescent women 14–18y; and menstruating adult women, both of which are positively skewed. The iron requirement distributions for children 1–3y, and 4–18y are also estimated by the US IOM (2001) to be skewed.
- Reliable data on the distribution of usual nutrient intakes for the group being assessed.
- Knowledge of the expected correlation between intakes and requirements among individuals: for all nutrients (except energy); this is assumed to be very low.

Reliable information on the distribution of usual intakes of the nutrient in the group can be obtained by adjusting the distribution of observed intakes statistically in an attempt to partially remove the effects of day-to-day variability in intakes (i.e., within-person variation) (Chapters 6). This yields an adjusted intake distribution, as discussed in detail in Chapter 3; Section 3.3.2. Such an adjustment can be applied, provided at least two independent days or three consecutive days of dietary intake data for a representative subsample of individuals in the group, have been collected. Four statistical methods are available for estimating the distribution of usual intakes. The methods have different features which are summarized in Box 3.3; Chapter 3; Section 3.3.2.

Figure 8b.3 shows an example of observed (unadjusted) and adjusted zinc intakes for a group of Malawian women. The intakes were obtained from two independent 24h recalls from each woman and adjusted to remove the effects of day-to-day (within-person) variability in intake following the procedure outlined in Chapter 3. Note that the adjustment process has yielded a distribution with the same mean (6.6mg/d) but reduced variability. Note that any bias arising from under- or over-reporting of food intakes is not removed by this adjustment process.

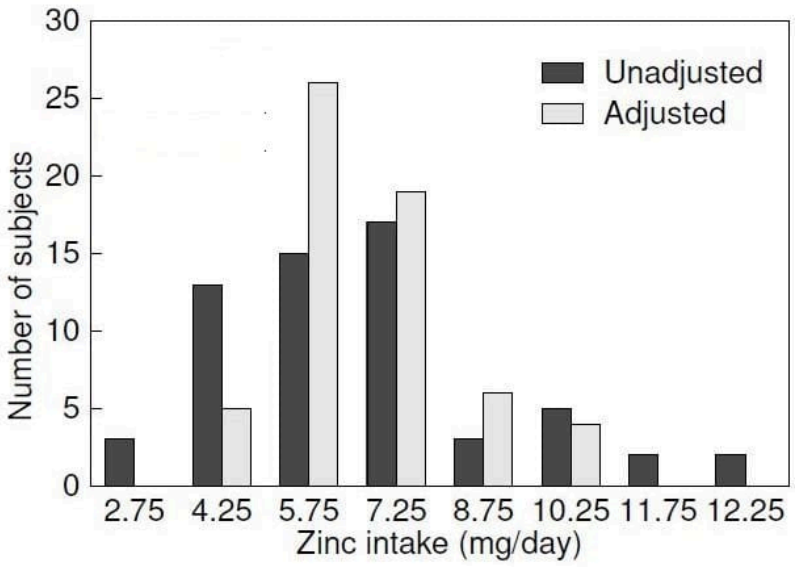


Figure 8b.3. Unadjusted and adjusted zinc intakes of some rural Malawian pregnant women. Data from Gibson and Ferguson (1998).

In this statistical approach, a probability of inadequacy can be calculated for any level of usual intake because information on the distribution of requirements for the group is known. Inadequate intakes associated with the usual intake of each individual in the group is determined first, followed by the average for the group as a whole. The latter is estimated as the weighted average of inadequate intakes for each individual, as outlined below (Barr et al. 2002).

The calculations for the probability approach can be performed manually or by a computer. For the manual calculation, first the nutrient intakes are classified into six classes defined by the AR and the associated SD limits as shown in Table 8b.5 (column A). The number of individuals with intakes of the nutrient within each class is then determined; this number is then multiplied by the appropriate probability for each class (Table 8b.5, column B) to give the number of individuals per class who were likely to have intakes below their own requirements. The numerical probabilities are derived from the area beneath the “normal” curve between the stated SD limits. The sum of the numbers in the six classes gives the total number of individuals in the population group with inadequate intakes for the nutrient. This sum can be expressed as a percentage of the total population group to give a prob-

ability estimate for the population group as a whole. For example, assume that in a population group of n = 4600, the numbers of individuals with usual intakes of vitamin C within classes one to six, respectively, are 300, 500, 800, 900, 1400, and 700. When multiplied by the appropriate probabilities for each class (Table 8b.5, column B), the number of individuals per class likely to have vitamin C intakes below their own requirements becomes 300, 465, 552, 279, 98, and 0. The sum of these numbers equals 1694, and this represents the total number of individuals with inadequate intakes for vitamin C. When expressed as a percentage, 37% of the total population are predicted to have intakes of vitamin C below their own requirements.

Table 8b.5. Assignment of "risk" or probability statements to six classes of observed intakes expressed as proportions of the Estima most nutrients is symmetrical with the coefficient of variation at 10% or 15%. Compiled from Beaton (1985).

Class	A. Individual's intake in terms of the distribution of requirements	B. Prol
1	< –2SD	1.00
2	–2SD to –1SD	0.93
3	–1SD to mean	0.69
4	mean to + 1SD	0.31
5	+ 1SD to + 2SD	0.07
6	> +2SD	0.00

The full probability approach must be used for assessing the prevalence of inadequate intakes of iron for women between the ages of menarche and menopause because the distribution of iron requirements for this life-stage group is highly skewed: some women have high menstrual iron losses. Calculations can be performed by using one of the four computer programs itemized in Box 3.3 Chapter 3; Section 3.3.2. Alternatively, for iron a spreadsheet can be used for these calculations in conjunction with tabulated data presented in Gibson and Ferguson (2008), provided the distribution of observed intakes have been adjusted statistically to yield information on the usual intakes. See: Gibson and Ferguson (2008). Tables in this publication provide data for the probability of inadequate intakes of iron (mg/d) for adolescent girls age 14–18y, adult women prior to menopause, and children age 1–3y and 4–8y. Different ranges of usual iron intakes are presented at three levels of bioavailability: low (5%), intermediate (10%); and high (15%). An example of how to use the table to estimate the prevalence of inadequate intakes of iron for menstruating women consuming a diet with 5 percent average bioavailability of iron is also given in Chapter 10 of Gibson and Ferguson (2008). Comparable tables for U.S. adolescent girls (14–18y) and menstruating women classified as non-oral contraceptive users or contraceptive users (assuming 60% reduction in menstrual iron loss) and based on 18% bioavailability are available in IOM (2001).

Note that the full probability approach should not be used for energy. Energy intake is highly correlated with requirement among non-obese individuals (even after age, sex, and weight adjustments have been made).

At present, the absence of reliable estimates of the ARs for all nutrients limits the general applicability of the probability approach to estimate the prevalence of inadequacy for every nutrient. Table 8b.6 summarizes the nutrients for which an AR and CV% of the requirement have been set by the U.S. Food and Nutrition Board (IOM, 2003). In the United Kingdom, ARs have been documented for iron, calcium, zinc, vitamin C, vitamin B_{12}, folate, riboflavin, and vitamin A (COMA, 1991). The WHO / FAO (2004) report also provides ARs for selected nutrients.

Table 8b.6. List of nutrients with dietary reference intakes (DRIs) for the United States and Canada. The √ indicates that the specific DRI has been defined for that nutrient. EAR, Estimated Average Requirement; RDA, Recommended Dietary Allowance; AI, adequate Intake; UL, Tolerable Upper Intake Level. From IOM (2000a); IOM (2000b); IOM (2001).

Nutrient	EAR	RDA	AI	UL
Calcium	√	√		√
Copper	√	√		√
Chromium			√	
Iodine	√	√		√
Iron	√	√		√
Magnesium	√	√		√

Nutrient	EAR	RDA	AI	UL
Manganese			√	√
Molybdenum	√	√		√
Phosphorus	√	√		√
Selenium	√	√		√
Zinc	√	√		√
Thiamin	√	√		
Riboflavin	√	√		
Niacin	√	√		√
Vitamin B_6	√	√		√
Folate	√	√		√
Vitamin B_{12}	√	√		
Vitamin C	√	√		√
Vitamin A	√	√		√
Vitamin E	√	√		√
Vitamin D	√	√		√
Fluoride			√	√
Sodium			√	√
Potassium			√	
Biotin			√	
Choline			√	√
Vitamin K			√	
Pantothenic Acid			√	

8b.3.2 EAR Cutpoint Method

A shortcut to the probability approach has been developed by Beaton (1994), and adopted by the U.S Food and Nutrition Board (IOM, 2000) for assessing the proportion of inadequate intakes in a group. This simpler version is termed the EAR cutpoint method and does not require information on the exact requirement distribution. The EAR cutpoint method can be used, providing the following three conditions are met:

- Intakes and requirements of the nutrient are independent (assumed to be true for all nutrients, but is not true for energy).
- The distribution of intakes in the population group must be more variable than the distribution of requirements (which is generally the case in free-living Individuals).
- The distribution of requirements in the group is approximately symmetrical about the AR (but this is not the case for the iron requirements of children aged 1 to 8y or women between the ages of menarche and menopause).

Again, data on the usual nutrient intakes are required, and these are best obtained by statistically adjusting the observed intakes for day-to-day variation (i.e., within-person variation), as discussed in Chapter 3. In this simplified version, the prevalence of inadequate intakes within the group is simply estimated by counting the number of individuals in the group with usual intakes below the AR, instead of estimating the level of inadequate intakes of each individual separately. This prevalence is represented by the shaded area to the left of the AR under the curve showing the distribution of usual intakes, as shown in Figure 8b.4. The larger area to the right of the AR represents the majority with usual intakes above the AR. Further theoretical justification for this approach can be found in Carriquiry (1999) and IOM (2003).

This EAR cutpoint method is especially useful when the actual prevalence of inadequate intakes in the group is close to 50%. As the true prevalence approaches zero, or 100%, the performance of this method declines, even if the conditions listed above are met. In addition, in circumstances when the variability in intakes is very low, as may occur for groups fed similar diets, then the full probability approach should be used instead of the EAR cutpoint method. Some examples of simulations used to assess the performance of the EAR cutpoint method in different situations are given in (IOM, 2000).

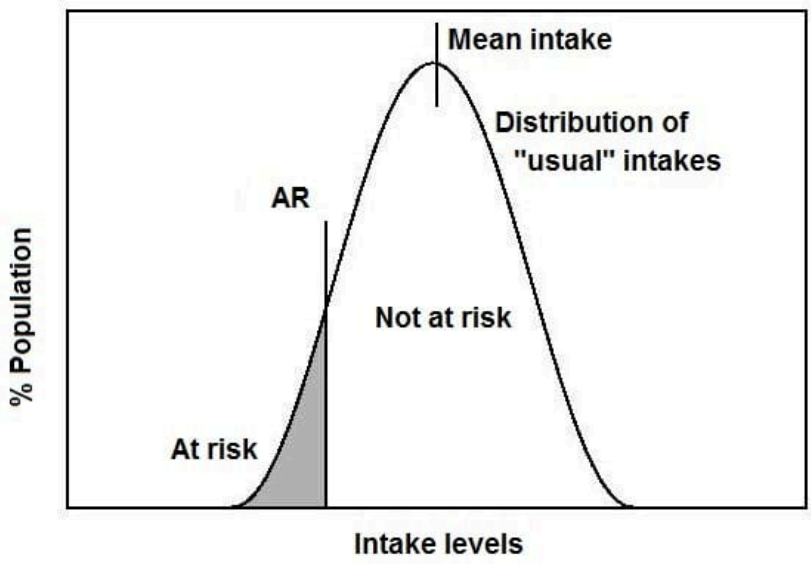


Figure 8b.4. The estimated average requirement cutpoint method for estimating the proportion of individuals with intakes below the AR.

Currently, the EAR cutpoint method based on the U.S Food and Nutrition Board Dietary Reference Intakes (NRVs equivalent) can be applied to assess the prevalence of inadequate intakes among population subgroups for the following nutrients: calcium, copper, iodine, magnesium, molybdenum, phosphorus, selenium, zinc, thiamin, riboflavin, niacin, vitamin B_6, folate, vitamin B_{12}, vitamin C, vitamin A, vitamin D, and vitamin E (Table 8b.6). As noted earlier the EAR cutpoint method cannot be used for assessing iron intakes for menstruating adolescents aged 14–18y and menstruating women because their distribution of iron requirements is skewed, although it can be used with these groups for the other nutrients listed above.

The EAR cutpoint method has been used to identify nutrients of concern when revising the USDA food assistance programs based on US national survey data (Trumbo et al., 2010), as well as for assessing the adequacy of usual zinc intakes from national food consumption surveys conducted in several low- and middle-income countries (Hess, 2017).

8b.3.3 Using 77% Percent of the Recommended Intake (RI_{98}) as a Cutoff Value

In some tables of NRVs, estimated average requirements (i.e, ARs) for nutrients are not specified, as noted earlier. In such cases, approximations for the estimated average requirements can be calculated, provided the RI (or equivalent) for each nutrient approximates the mean requirement estimate plus two standard deviations, with a specified coefficient of variation (i.e., RI_{98}) (NRC, 1986). The proportion of the population with usual intakes below the derived AR is then calculated, as described in Section 8b.3.2.

A cutoff value of 77% of the RI is sometimes used. This approach assumes a CV for the nutrient of 15% about the AR. Such an assumption will yield a conservative estimate of nutrient inadequacy compared to that based on a CV for the nutrient of 10% about the AR. Use of the latter CV corresponds to a cutoff of approximately 83% of the RI instead of 77%; hence, it will yield a larger percentage of the group likely to have inadequate intakes. Briefel et al. (2000) used 77% of the 1989 U.S RDA values for zinc as a cutoff to evaluate the zinc intakes of the U.S population in the National Health and Nutrition Examination Survey III 1988–1994, prior to the release of the EAR for zinc by the IOM (2001).

Few ARs have been published by WHO / FAO. However, a table of conversion factors that can be used for calculating ARs from the WHO / FAO (2004) RNIs (equivalent to RIs) is available in WHO (2006) (Annex C) along with a table in the text depicting the calculated AR values. No ARs can be calculated from the WHO / FAO RNIs for iron for adolescent girls, menstruating women (19–50y), or children (< 9y) because of the skewed distribution of requirements for iron for these groups. The conversion factors that have been used in WHO (2006) to calculate the WHO/FAO ARs are based on standard deviations derived by the U.S Food and Nutrition Board of the Institute of Medicine and used to calculate the U.S and Canadian Recommended Dietary Allowances (RI equivalent).

8b.3.4 Use of Adequate Intake

The Adequate Intake (AI) is estimated in a number of different ways, depending on the nutrient and the life stage group. When the AI has been set based on the mean intake of a healthy population, then if the study group's mean intake is at or above the AI, the group can be assumed to have a low prevalence of inadequate intakes (Table 8b.1), although the actual prevalence of inadequate intakes cannot be established. When the AI is used in this way, the distribution of intakes does not have to be adjusted for within-person variation because only the mean or median intake is being examined for comparison. However, no assumptions can be made about the prevalence of inadequate intakes when the mean or median intake of a group falls below the AI because the distribution of requirements is unknown (Barr et al. 2002; Murphy and Poos, 2002).

8b.3.5 Use of Upper Intake Level

When any comparison is made with the Upper Intake Level (UL), it is important to note if the UL assessment for the nutrient of interest is based on information on usual daily intakes from all sources or only from supplements, fortificants, and medications. The prevalence of potentially excessive intakes is estimated by assessing the proportion of the group with usual intakes above the UL. The goal is to have a low proportion of a group with intakes above the UL (Barr et al. 2002; Murphy and Poos, 2002). Before an assessment is made, the intake distribution must be adjusted to remove the effect of within-person variation in intakes (see Chapter 3). Because of the uncertainties and the variability in individual sensitivities to the adverse effects, the actual prevalence of adverse health effects in the group will seldom be equal to the proportion with intakes above the UL. Moshfegh et al.(2005) assessed the proportion of the population in the U.S NHANES (2001–2002) survey with intakes above the UL for eleven nutrients.

Table 8b.7 provides an example of the use of all the approaches described above for assessing nutrient intakes of a group derived from a national nutrition survey. Further details are given in Murphy and Poos (2002).

Table 8b.7. Data and evaluation of a hypothetical group's diet. Data for children 4–8y, from 1994–1996 Continuing Survey of Food Intakes by Individuals. UL, Upper Intake Level; AR, Average Requirement; AI, adequate Intake; RI, Recommended Intake; inadequacy, percentage of intakes below the AR; Excessive intakes, percentage of intakes above the UL. The intakes were adjusted to remove the effects of day-to-day variation before these assessments were made. Note that the AI for calcium was replaced by an AR in 2011, so the appropriate assessment method now would be to determine the proportion of the group with intakes below the AR. From Murphy and Poos (2002).

	Thiamin (mg)	Magnesium (mg)	Calcium (mg)	Vitamin C (mg)
Group Mean intake	1.44	212	838	96
Requirement	0.5 (AR)	110 (AR)	800 (1997 AI)	22 (AR)
inadequacy (%)	< 1	5	Low	< 1%
UL	None set	110	2500	650
Excessive intake (%)	Unknown	Unknown (supps only)	< 1	< 1

8b.3.6 Energy Intakes

As noted for individuals (Section 8b.2.6), the probability approach cannot be used to estimate the prevalence of inadequate or excessive intakes of energy in a group because of the expected high correlation between energy intake and energy requirements

among individuals. The latter violates one of the assumptions required for the adoption of the probability approach. Instead, the group mean energy intake can be compared with the mean estimated energy requirement (EER) for the group, taking into account the ages, sexes, heights, weights, and physical activity levels of the individuals in the group. If the mean intake is equal to the mean EER, energy intake can assumed to be adequate, whereas if the mean intake exceeds or falls below the EER, then the group on average is likely to be gaining or losing weight, respectively. Again, because only the mean is calculated, the intake distribution does not need to be adjusted for within-person variation in intakes, as noted for the AI (Section 8b.3.4).

Alternatively, as under-reporting of energy intake in dietary assessment is ubiquitous, the adequacy of the group's energy intake can also be assessed using the body mass index (BMI). The proportion of a group of U.S adults with a BMI below or above the normal range of18.5–25kg/m^2 would be classified as having inadequate or excessive energy intakes, respectively for their activity level, whereas for U.S children and adolescents, the proportion with BMI-for-age < 5th percentile or ≥ 85th percentile should be used (Kuczmarski and Flegal, 2000).

For international use for children 0-5y, underweight and overweight are defined as BMI-for-age <–2 Z-score and above +2 Z-score, respectively, of the WHO Child Growth Standards 0–5y (WHO, 2006), but <–2 Z-score and above +1 Z-score for the BMI-for-age WHO Growth Reference for children 5–19y (de Onis et al, 2006).

8b.3.7 Macronutrient Intakes

For macronutrients with a reference intake range, the proportion of the group with usual intakes (adjusted for within-person day-to-day variation) outside the AMDR can be assessed. If the prevalence outside the range is high, then an intervention may be needed.

This page titled 6.3: Using NRVs to assess nutrient intakes of groups (8b.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

6.4: Using the NRVs for planning nutritious diets (8b.4)

As shown in Figure 8b.1, NRVs can also be used to plan diets at the individual or group level. Planning for an individual involves setting intake targets that ensure a nutritionally adequate diet, while also being acceptable, whereas for a group, the goal is to determine a desirable intake distribution that minimizes the prevalence of inadequate or excessive intakes. Numerous applications for planning nutritious diets exist, including dietary counseling for individuals, developing menus for institutions, or designing interventions for high-risk groups, as shown in Table 8b.2.

8b.4.1 Using the NRVs for planning nutritious diets of individuals

The goal of planning a diet for an individual is to ensure there is a low probability of inadequacy and the potential risk of an excessive intake is small. This can be achieved by ensuring that the usual intake of the individual meets the RI_{98}, because by definition, the intake at that level should be adequate for almost all (98%) individuals. Use of the AR as a goal for the nutrient intake of an individual is not appropriate because an intake at this level would be adequate for only 50% of individuals. When setting the RI_{98}, however, knowledge about the distribution of requirements for many nutrients remains limited, so an assumption about the coefficient of variation of the requirement is often made, as noted in Chapter 8a. This has led to debate over the use of the RI_{98} as the goal for planning diets at the individual level (Beaton, 2006; Murphy et al. , 2006). Nevertheless, at present the RI_{98}remains the most appropriate target for the nutrient intake for individuals, at least until more accurate data on the requirement distributions are available.

To avoid a potential risk of adverse effects, the nutrient intakes of an individual should not exceed the UL. This means that the diet of an individual should have usual nutrient intakes that fall between the RI_{98} and the UL. For those nutrients with only an AI, then nutrient intakes for an individual should fall between the AI and the UL.

Planning for an individual's energy intake is unlike other nutrients. For most nutrients, there is essentially no risk if an individual consumes an intake that may marginally exceed his or her requirement (i.e., the RI_{98} exceeds the requirements of almost all (97–98%) individuals in a group). However, in the case of energy, consuming an energy intake that met the requirements of 97–98% of a group would lead to weight gain in that proportion of individuals. Accordingly, energy requirements of individuals are best monitored by assessing their body weight status — those with BMI in the normal range who are maintaining their weight (or gaining weight appropriately during growth) are considered to be meeting their energy requirement. If an estimate of energy needs is required for planning purposes, an approximation can be provided by using the equations for EER, which consider age, sex, height, weight and physical activity level. However, because there is variability in energy requirements even among individuals with the same characteristics, body weight must be monitored and intake adjusted where necessary (Barr et al., 2003). For macronutrients (expressed as percentage of energy), the goal is to ensure that the intakes from the individual's diet falls within the AMDR ranges.

Table 8b.8. Data and assessment of the nutrient intakes of a hypothetical individual's diet. Data for a 30y man observed for 3 days and thus assumed to reflect the usual intake. DFE: Dietary folate equivalent. RI_{98} recommended intake. UL Upper Intake Level. From Murphy and Vorster (2007).

	Thiamin (mg)	Ribo-flavin (mg)	Folate (µg) DFE	Zinc (mg)	Phos-phorus (mg)
Man's avg. daily intake	1.3	1.1	400	10.3	3,800
RI_{98}	1.2	1.3	400	11.0	700
UL	None	None	1,000	40	4,000
Proposed change in in intake	None	Inc-rease	None	Inc-rease slightly	None

Planning a nutritious diet at the individual level may involve counseling individuals with special needs such as vegetarians and athletes, or those requiring therapeutic diets, as well as providing food-based dietary guidance. Table 8b.8 provides an example of how to plan to change the nutrient intakes of a hypothetical 30y male individual. In this example, provided the three days of intake collected represented his usual intake, the individual has been advised to increase his intake of both riboflavin and zinc so that they meet the RI_{98}; no other changes were needed as none of the intakes exceeded the UL (Table 8b.8). Note that in this example, the evaluation of the individual's nutrient intake was conducted using the more practical qualitative approach described in 8b.2.4, rather than the statistical approaches described in 8b.2.1 through 8b.2.3.

8b.4.2 Using NRVs for planning nutritious diets for groups

Planning diets for groups is more complex than for individuals because of the wide range of nutrient intakes in a group arising from differences in the amount and types of food consumed. The goal is to alter the distribution of current nutrient intakes within the group to reduce the prevalence of inadequate and excessive intakes, where necessary. The three NRVs used for planning nutrient intakes for a group are the AR, the AI, and the UL; the RI_{98} should not be used for groups. For more details on planning diets for groups, the reader is referred to IOM (2003) and Murphy and Barr (2005).

The EAR cutpoint method described in Section 8b.3.2 is also used for planning the diets of groups, with the goal of minimizing the proportion of the group with intakes below the AR and above the UL. Steps in planning diets for groups, using the EAR cutpoint method are shown in Box 8b.2. The method can only be used for those nutrients with a defined AR and UL, and for which the requirement distribution is approximately symmetrical.

Box 8b.2. The four steps involved in planning diets for groups using the EAR cutpoint method

- Decide on the goals of the planning process. This involves deciding on what is considered an acceptable, practical, and low prevalence of undesirable intakes (i.e., inadequate intakes below the AR and excessive intakes above the UL) given the situation. The acceptable prevalence of inadequacy is often 2–3%, although this may not always be practical.
- Select the target distribution of usual intakes that meets the defined goals. This means that knowledge of the current intake distributions for the nutrients of concern is already available. A change may be required for only certain nutrients to either increase the intakes (i.e., move the current distribution to the right) so that fewer people have intakes for those nutrients below the AR, or decrease intakes (move the current distribution to the left) so fewer people have intakes above the UL.
- Plan a menu to achieve the goals for all of the nutrients of concern. To accomplish this step, first the nutrient content of the menus must be calculated using appropriate dietary assessment software and food composition data, and taking into account the foods and amounts that will be actually consumed rather than those provided by the menus. The goal here might be for the menu to provide nutrient levels at the midpoint of the target distribution for each nutrient of concern.
- Assess the results following the implementation of the new menu. This stage involves collecting dietary data from the group of interest, calculating their nutrient intakes, and comparing the usual intake distributions for the nutrients of interest to the goals set in step 1. If the planning goals have not been met, then some or all of the steps may need to be repeated until the results are satisfactory.

Modified from Murphy and Barr (2005).

As noted earlier, because iron requirements for certain life-stage groups are skewed, the full probability approach must be used to calculate the prevalence of inadequate or excessive intakes for these groups using software described earlier (Chapter 3; Section 3.3.2). Alternatively, published tables of the distribution of iron requirements can be used to estimate manually the probability of inadequacy for each individual within the group, from which the probability of inadequacy for the whole group can be calculated; refer to Section 8b.3.1 for more details.

Note that irrespective of which method is used, the distributions of observed intakes, both before and after the intervention, must always be adjusted statistically to yield the usual intake distribution and thus remove the effects of day-to-day variations; see Chapter 3 for details of the statistical techniques to use.

When only an AI has been set for a nutrient, it is not possible to estimate the prevalence of inadequate intakes in the group for that nutrient, and instead the goal is to ensure that the average intake of the group for that nutrient is at or above the AI.

When planning the energy intakes of groups, the goal is for the mean intake of the group to be approximately equal to the EER. The latter represents the average energy requirement for the group. No RI_{98} is set for energy because intakes above requirements lead to weight gain. The EAR cutpoint method should not be used for planning energy intakes (Murphy and Barr, 2005). Because only the mean intake for the group is calculated, the intake distribution does not need to be adjusted for day-to-day variation, as noted earlier.

When considering the AMDR, the diets of groups should be planned to minimize the prevalence of intakes outside the AMDR. In this way, the risk of chronic diseases will be reduced, while ensuring adequate intakes of other essential nutrients (Murphy and Barr, 2005).

Example of the process for planning zinc intakes using the EAR cut-point method

In the example shown in Table 8b.9, the first step is to select the planning goal, which in this example will be to have no more than 2–3% of girls aged 9–13y to have usual zinc intakes less than the AR (i.e., 7mg) and no more than 2–3% above the UL (i.e., 23mg/d). Step 2 is to estimate the target intake distribution. To complete this step, information on the current intake distribution for zinc is needed. In this example, data on the usual zinc intake distribution of girls aged 9–13y from NHANES III is used (Table 8b.9). Note that the estimated prevalence of inadequate intakes based on this current intake distribution is 9%, above the goal of no more than 2–3% below the AR. Note also that the AR for zinc of 7mg approximates the 10th percentile of the current usual intake distribution. Hence, assuming our planning goal is to reduce the prevalence of inadequate intakes to 2%, then the distribution of usual intake needs to be shifted such that the current intake at the 2nd percentile (currently 6.2mg/d) is increased to correspond to the AR of 7mg. Thus, an increase of 0.8mg is needed, and assuming that the shape of the distribution does not change, then intake at every percentile would also increase by 0.8mg/d. This would yield a target median intake of 9.4 + 0.8 = 10.2mg as shown in Table 8b.9. Figure 8b.5 shows the current usual intake distribution and the proposed target usual intake distribution for this group of girls aged9–13y.

The next step is to plan a menu to meet the target intake distribution, which in this example could be achieved by substituting some of the foods currently consumed with those higher in zinc, recognizing that not all food will be consumed. In the final step, the effect of the new menu is examined. This can be achieved by collecting dietary intake data from each girl in the group, calculating zinc intakes, and after adjusting the distribution of observed intakes to usual zinc intakes, estimating the prevalence of inadequate and possible excessive intakes. If the planning goal of 2% with intakes below the AR has not been met, then a new plan will be required.

Table 8b.9. Current and target usual zinc intake distributions (mg/d from food) for girls 9–13y.
** NHANES III data. The distribution is for food intake only, assuming that the girls for which planning is being done do not take supplements.*
† AR = 7mg/d, ‡ UL = 23mg/d
From Murphy and Barr (2005).

	Current Intake*	Target Intake	Change
Mean	9.6	10.4	+0.8
Percentile			
2nd	6.2	7.0	+0.8
5th	6.5	7.3	+0.8
10th	7.1	7.9	+0.8
25th	8.1	8.9	+0.8
50th	9.4	10.2	+0.8
75th	10.9	11.7	+0.8
90th	12.5	13.3	+0.8
95th	13.5	14.3	+0.8

	Current Intake*	Target Intake	Change
99th	15.5	16.3	+0.8
% below EAR†	9%	2%	–7%
% above UL‡	0%	0%	None

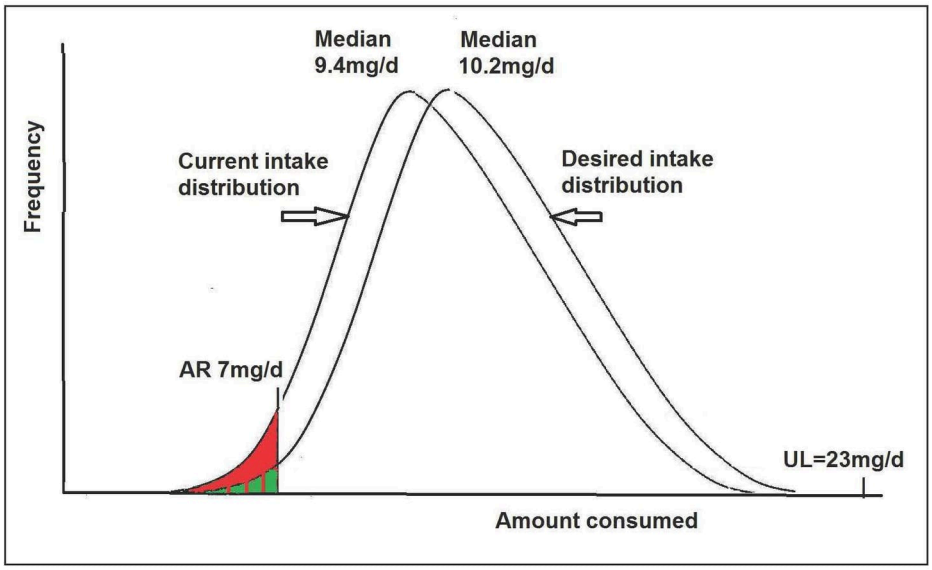


Figure 8b.5. The current distribution of usual intakes of girls aged 9–13y for whom the prevalence of inadequate intakes (% < AR) is about 9% (shaded red). Also shown is the desired/target distribution of usual intakes. Note that this improved distribution is shifted upwards so that the prevalence of inadequate intakes (% < AR) is now only 2% (shaded green).

In this example only zinc is considered whereas in practice intakes of all nutrients would be considered. The planner then has the task of deciding which of the nutrient intakes need to change necessitating the use of a nutrient calculation software package. The nutrient profiles of the proposed menus can be calculated using the software, so that the necessity of further changes can be assessed.

Health Canada followed the process outlined above when planning diets for their publication entitled “Eating Well with Canada's Food Guide” (Health Canada, 2011). Their goal was to ensure a low prevalence of nutrient intakes below the AR when a life-stage and sex group chooses foods from food groups following the recommended dietary pattern; see Katamay et al. (2007) for more details of the methods applied. However, the process to develop the subsequent Canadian Food Guide, released in 2019, did not include a comparable assessment of the nutrient adequacy of the recommended dietary pattern (Barr, 2019). A modification of this approach has also been used in the U.S to identify nutrient targets for school meals(IOM, 2010b).

There are several other challenges that may be encountered when planning diets for groups. These are summarized in Murphy and Barr (2005) and discussed in more detail in IOM (2003). As an example, all the approaches described above for planning diets for groups assume that the age and sex of the individuals within the group are similar, thus permitting a single set of ARs and ULs to be applied. However, in some circumstances, groups are heterogeneous making the planning process more challenging. Currently the recommended approach for heterogeneous groups is to first identify the subgroup with the greatest nutrient requirement per 1,000kcal (i.e., nutrient density) and second, apply the approaches described above to plan the intakes for each nutrient. Here care must be taken to ensure that no subgroups within the group have the potential for excessive intakes. Finally, after the new menus or food patterns have been implemented, then the actual nutrient intake distributions must be checked to ensure that the prevalence of inadequate and potentially excessive intakes remains low.

Applications for groups include planning diets for institutional feeding such as school lunch programs, residential schools, hospitals, prisons, nursing homes, the military etc, and in some cases, emergency food assistance programs (Murphy and Barr, 2005). The approach is also used for establishing appropriate levels of nutrient fortification of foods and described in detail in WHO (2006). For more details of these applications, the reader is advised to consult Murphy and Vorster (2007).

This page titled 6.4: Using the NRVs for planning nutritious diets (8b.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7: DIETARY GUIDELINES AND QUALITY (CHAPTER 8C)

Abstract

This chapter covers two related but distinct topics: Food-based dietary guidelines and assessment of population-level diet quality. Definitions of diet quality, food-based dietary guidelines, and assessment measures have all evolved over time, reflecting new and emerging scientific evidence for diet-health relationships.

Food-based dietary guidelines are developed by national health authorities for use by consumers, practitioners, and policymakers. Such guidelines translate current scientific knowledge about relationships between food intakes and health outcomes into clear population-level guidance. In this chapter, we provide a brief overview of the development and communication of guidelines and describe a global repository for food-based dietary guidelines. We summarize areas of consistency across guidelines and highlight some possible future directions, including incorporation of environmental sustainability concerns.

In the second part of the chapter, we describe approaches to assessing diet quality, particularly at the population level. Numerous metrics (indicators, indices, and scores) have been developed to summarize diet quality, reflecting different definitions of diet quality and also different measurement objectives. Widely used metrics include measures of adherence to national food-based dietary guidelines such as the U.S. Healthy Eating Index, measures of adherence to healthy regional diets such as Mediterranean diet scores and indices, and other measures developed based on epidemiological evidence linking dietary patterns and characteristics to health outcomes.

Using a series of examples, we describe how diet quality metrics are developed, evaluated, and used. Most scores and indices require detailed quantitative dietary intake and food composition data. However, several measures have been developed for use in contexts where such data are not available and simpler measures are required, and we include examples of these. Finally, as with food-based dietary guidelines, concerns with environmental impacts of diets and food systems are growing, which will likely result in new metrics incorporating this dimension.

This chapter covers two related but distinct topics: Food-based dietary guidelines, often provided by national health authorities, and assessment of diet quality. The section on assessment of diet quality has a primary focus on applications at the population level, for monitoring diet quality, assessing diet-disease associations, and evaluating policy and programmatic interventions. The two topics in this section are linked in this population-level focus, and because assessment of diet quality is sometimes approached through assessing adherence to national food-based dietary guidelines. However, the section also covers a range of other approaches to assessing diet quality in populations. Learning objectives for readers include gaining understanding of:

- The role of food-based dietary guidelines in communicating guidance from national health authorities;
- The range of topics covered by food-based dietary guidelines, and some areas of consistency and difference across various national guidelines;
- The evolution of definitions of diet quality, and how this has influenced assessment of diet quality;
- The wide diversity of diet quality indices currently in use;
- Some of the main types of indices, including those based on national food-based dietary guidelines, on regional diets, and on other summaries of evidence for diet-health relationships;
- How indices differ in their development, measurement (data sources), calculation and use;
- Which indices have lower data requirements for their calculation;
- How new concerns with sustainability and planetary health may affect diet quality definitions and assessment.

7.1: Food-based dietary guidelines (8c.1)

7.2: Assessing diet quality (8c.2)

7.3: Examples of diet quality indices (8c.3)

7.4: Alternate Healthy Eating Index (8c.4)

7.5: Mediterranean Diet Score or Index (8c.5)

7.6: Dietary Approaches to Stop Hypertension scores
7.7: Healthy Diet Indicator (8c.7)
7.8: Dietary Inflammatory Indices (8c.8)
7.9: Diet quality based on level of food processing (8c.9)
7.10: Lower-burden diet quality indices (8c.10)
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)
7.12: Summary of diet quality indices (8c.12)

This page titled 7: Dietary guidelines and quality (Chapter 8c) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Food-based dietary guidelines (8c.1)

This section provides a brief overview of the development and communication of food-based dietary guidelines (FBDGs), with a focus on those developed by national authorities. The section also describes a global archive for FBDGs and comments on areas of consistency across national guidelines. The section concludes with comments on likely future directions for FBDGs.

Science-based guidance on healthy diets is available from a variety of sources, including from national governments, international authorities such as the World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO), and medical and dietetic professional societies, among others.

FBDGs translate current scientific knowledge about relationships between food intakes and health outcomes into clear population-level guidance. They are "food-based" in contrast to guidance on nutrient intakes such as various Nutrient Reference Values (see Section 8b.1). They often provide advice and guidance on dietary patterns — what to eat and drink, how often, and in what amount. In this Section, we focus on FBDGs developed at the national level.

The earliest dietary guidance from national governments was focused on ensuring nutrient adequacy of diets (Harper, 1987). However, by the late 20th century the focus shifted to include a strong emphasis on reducing risks for non-communicable diseases (NCDs, such as cardiovascular diseases, metabolic syndrome, type 2 diabetes, and cancers), while still ensuring nutrient adequacy.

More recently, there have been calls to integrate considerations of environmental sustainability into FBDGs; however, to date this has been rare(Gonzalez Fischer & Garnett, 2016; Springmann et al., 2020). Some FBDGs also integrate guidance on physical activity, food safety, food skills (such as choosing healthier foods while shopping, and preparing meals), the role of ultra-processed foods, and other topics.

8c.1.1 Development of national-level food-based dietary guidelines

WHO and FAO have provided guidance to countries on how to develop FBDGs (WHO/FAO, 1998). This development process for national FBDGs was summarized by the European Food Safety Authority (European Food Safety Authority (EFSA), 2010) as including the following steps:

- Identify diet-health relationships (based on national but also global evidence);
- Using national data sources, identify:
  - Diet-related health problems;
  - Nutrients of public health importance;
  - Food groups relevant for FBDGs (for example, food groups with known links to health outcomes, and for sources of nutrients of public health importance);
  - Food consumption patterns;
- Test FBDGs with audiences and refine them;
- Develop graphical representations of FBDGs.

The development of FBDGs requires synthesis of a wide range of evidence and contextual information, although in practice, some types of information may not be available. For example, many countries lack recent and nationally representative information on food consumption patterns.

The process also calls for the involvement of multiple stakeholders and can be both a technical and a political process. Through considering food consumption patterns and cultural dietary preferences, the resulting guidelines should allow for flexibility in food choices, addressing socioeconomic constraints and respecting cultural diversity within national populations (WHO/FAO 1998).

A majority of national FBDGs include some type of summary graphic, sometimes called a food guide, to convey concepts like diversity and proportionality among different food groups. The graphics sometimes also include specific details on quantities (standard servings per day or per week) for various food groups.

Examples of summary graphics include images of healthy plates, food pyramids, food baskets, or other cultural shapes Three examples are shown in Figure 8c.1 — Malta, Oman, and Estonia. (Altamirano Martínez et al., 2015; van't Erve et al., 2017; Herforth et al., 2019). Some countries provide several graphic images, such as both a pyramid and a healthy plate. See, for example, images

for the Finnish FBDGs and the Swiss FBDGs.



Malta



Oman



Estonia

Figure 8c.1: Examples of graphic images from national food guides from Malta, Oman and Estonia. Reproduced with permission from FAO.

8c.1.2 Communication and implementation of national-level food-based dietary guidelines

FBDGs aim to provide actionable guidance for consumers and for public health and medical practitioners. Full implementation of FBDGs includes not only development of a variety of targeted consumer education strategies and materials, but also integration within a range of sectoral policies and programs to help ensure that healthy diets are available, accessible, and affordable (Wijesinha-Bettoni et al., 2021). Policies and programs include those related to public procurement of foods (for example, for feeding in schools and other institutions), taxes and food subsidies, and various food security, nutrition and agricultural policies and programs. More information on implementation of FBDGs is available from FAO.

Given these varied uses, governments develop a variety of documents and tools to communicate and implement the FBDGs. These may include:

- Technical documents detailing the process of evidence review and justifying each guideline statement, or key message. Technical documents may include results of systematic reviews of evidence on diet-health relationships, diet modeling studies, and/or references to systematic reviews from other sources (see, for example,New Zealand's documentation of evidence);
- Sector-specific policies, regulations and guidance; for example, policies and guidance for implementation of school meals programs that incorporate the national FBDGs (see, for example, U.S. regulations for school meals programs);
- Guidance for public health and medical practitioners, which may be at an intermediate technical level, and with a practical focus to help them translate the FBDGs for consumers (see, for example, Canada's Dietary Guidelines for Health Professionals and Policy Makers);
- Short, simple consumer-facing documents such as brochures, posters, and brief graphical food guides with advice on food groups (see, for example, the food guide graphic and tri-fold brochure from Benin);
- Consumer-facing internet pages and applications that provide a variety of tools for learning and self-assessment (see, for example, Canada's Food Guide webpage).

When evaluating the impact of programs and policies, it is often desirable to measure and assess adherence to dietary guidance. See Section 8c.2 for a discussion of different approaches to assessing diet quality and adherence to dietary guidance. Once fully implemented, the FBDGs themselves should also be evaluated for their effectivenes; information on evaluation of FBDGs is also available at FAO. However, up to the present, information on effectiveness is limited.

8c.1.3 Repository of national food-based dietary guidelines

The FAO maintains a repository of national FBDGs, which currently (December, 2021) includes FBDGs for 95 countries. Each country page includes standard information provided by national authorities for most or all countries (see Box 8c.1).

Box 8c.1 Information no national food-based dietary guidelines.

- Official name of the country's dietary guidelines;
- Publication year of the most recent guidelines;
- Development process and stakeholders involved;
- Intended audience and age groups covered;
- Description of the food guide graphic;
- Key messages;
- Links to a variety of downloadable materials, including food guides and technical documents, and to relevant government websites if applicable.

In addition to the information in Box 8c.1, available for most or all countries, newly updated country pages may also include information on implementation, evaluation, and incorporation of sustainability concerns.

FBDGs are currently (December, 2021) available at the repository for 33 European countries, 31 countries in the Americas, and 18 countries in the Asia-Pacific region, but for only 7 African countries. Availability varies by country income classification, with FBDGs available for slightly over half of all high- and upper-middle-income countries, just over one-quarter of lower-middle-income countries, and for fewer than one in ten low-income countries (based on World Bank 2021 country income classifications).

8c.1.4 Coverage of sub-populations

Countries have taken different approaches to providing guidance for specific age groups and physiological conditions:

- Some countries have taken a life-cycle approach in a single dietary guidelines document, and include sections with specific guidance for infants, young children, school-aged children, adolescents, and for pregnant and lactating women (see, for example, Kenya's National Guidelines for Healthy Diets and Physical Activity).
- Other countries have developed separate technical documents and/or consumer-facing guidance documents for some or all age and physiological groups (see, for example, New Zealand's technical background papers and consumer resources for different age groups).
- Other countries state that the main dietary guidelines document and key messages apply to all healthy individuals ≥ 2y, but they may still highlight the needs of age groups or other sub-groups (vegetarians, pregnant or lactating women) within the

main guidelines document (see, for example, Lebanon's Food-based Dietary Guidelines).
- The most common age group with separate and specific guidance is infants and young children under two years of age (see, for example, technical and consumer-facing FBDG for Panama).

8c.1.5 Consistency of national food-based dietary guidelines

Commonalities and differences among national FBDGs graphics and key messages were described in several recent reviews (Altamirano Martínez et al., 2015; van't Erve et al., 2017; Herforth et al., 2019). All reviews summarized food guide graphics and noted that the two most commonly used graphics were versions of a pyramid or a plate / food circle. The remainder of this section is based on the review by Herforth et al. (2019).

Nearly all food guide graphics conveyed the importance of food group diversity and proportionality — that is, that some food groups should be consumed in larger quantities than others. However, only about one-third gave quantitative information on recommended amounts per day or per week (either in grams or as the number of servings). Most graphics included depiction of fats and oils and of sugars / sweets, showing them as small in proportion to other groups. About one-third included depictions of other foods to be consumed in moderation, such as pizza, hamburgers, or salty snacks. About half included depictions of water, and about half included depictions of physical activity.

Excluding depiction of food groups to consume in moderation, the most common number of food groups depicted was four or five, with the most common groupings being: (1) Starchy staple foods; (2) Fruits and vegetables (either as one group or as two separate groups); (3) Dairy; and (4) Other protein foods (meat, poultry, fish, eggs, legumes, and sometimes nuts).

However, there was variability in whether legumes were grouped with protein foods, with vegetables, or with both. Dairy was frequently a separate group but was sometimes grouped with other protein foods. Many countries did not depict nuts or grouped them with fats and oils rather than with protein foods. Other inconsistencies related to how potatoes were classified (as starchy staples, or as vegetables) and in whether fruit juice was depicted with whole fruits.

Herforth et al. (2019) analyzed key messages in addition to graphics. The most consistent key messages were aligned with quantitative global guidance from the WHO, including messages on ample consumption of fruits and vegetables and moderation in intake of salt, added sugar, and fats (particularly saturated and trans-fats). Only about one-half of countries emphasized whole grain intake. Key messages encouraging dairy intake were nearly universal in Europe and North America but were less common in Asia, Africa, and Latin America. Some FBDGs provided inconsistent messages on unsaturated fats, with positive key messages indicating to choose them but with graphics grouping the healthier choices (nuts and oils) together with less healthy ones (saturated fats) as groups to moderate.

8c.1.6 Future directions for food-based dietary guidelines

Recent FBDGs from several countries have highlighted the need to consider environmental sustainability in addition to concerns with nutrient adequacy and NCD risk reduction. For example, Brazil (2014), Uruguay (2016), Ecuador (2018), and Sweden (2015) all identify environmental sustainability as a guiding principle or central focus for their FBDGs, and Qatar (2015) also includes a key message on environmental protection. WHO and FAO have also articulated guiding principles for sustainable healthy diets (FAO & WHO, 2019). As noted above, the FAO repository for FBDG has begun documenting country-level attention to sustainability.

Several recently published FBDGs also include a stronger focus on behaviors to help achieve healthy eating patterns. Several of the same countries noted above (Brazil, Ecuador, Uruguay), as well as Canada, include key messages and guidance around food skills, social aspects of eating (enjoying meals with family and with others), and/or being an informed consumer through reading labels and understanding food industry marketing practices. A number of countries also provide guidance to reduce intake of highly processed foods.

In summary, future FBDGs will likely continue to focus on nutrient adequacy and NCD risk reduction, but may move toward incorporating new themes of sustainability, food skills, and messaging around highly processed foods. In addition, increased attention to go beyond careful development of FBDGs to thorough implementation and evaluation is also warranted (Wijesinha-Bettoni et al., 2021). Finally, FBDGs will continue to be updated to reflect new knowledge as well as changes in lifestyles and food supplies.

This page titled 7.1: Food-based dietary guidelines (8c.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.2: Assessing diet quality (8c.2)

This section starts with brief comments on definitions of diet quality, and how these have changed over time. We then provide detailed discussion of the various ways researchers have measured diet quality, both in high-income "data-rich" settings, and in more resource-constrained settings. As noted earlier, our focus is on measurement for population-level applications. Numerous metrics (indicators, indices, and scores) have been developed to measure the quality of the diet. See Box 8c.2 for comments on vocabulary for metrics.

Box 8c.2 Metric, indicators, indices and scores.

"Metric" is a general term meaning a standard against which something is evaluated. Other similar terms include "indicator" or "index" — indicators are sometimes thought of as simple and reflecting only one dimension, and indices as composites summarizing across multiple dimensions. However, these terms are used differently by different authors. Metrics discussed in this section may be called indicators, indices, or scores. We use the name used by the authors who first developed or reported on the metric. Since many of the most common diet quality metrics are called an "index", we generally use "index" or "indices".

Some indices are based on national FBDGs but others are not, and have a different rationale. For recent reviews of diet quality indices, see: Burggraf et al., 2018; Trijsburg et al., 2019; Aljuraiban et al., 2020; and Miller et al., 2020.

The section covers:

- How the definition of diet quality has changed over time
- The basis, or rationale for various diet quality indices
- Uses of diet quality indices
- The development process for diet quality indices
- Detailed examples of different types of diet quality indices

By providing detailed examples of diet quality indices, we aim to illustrate their variety and uses. The section closes with a brief discussion of new efforts to incorporate the idea of environmental sustainability into diet quality definitions and indices.

8c.2.1 Changing definitions of diet quality

Diet quality indices have evolved along with changing definitions of diet quality (see Box 8c.3). Early indices such as the "Mean Adequacy Ratio"(Guthrie et al., 1972) aimed to summarize nutrient adequacy only. Newer diet quality indices often aim to reflect multiple characteristics of diets. Many indices aim to reflect both nutrient adequacy and characteristics of diets related to NCD risk. Currently, efforts are underway to also incorporate sustainability concerns.

Box 8c.3 Evolution of diet quality definitions

Historically, ideas about diet quality were a response to deprivation, whether due to poverty or the demands of wartime economies. Healthy and high-quality diets were defined as those that were sufficient: first in calories and protein, and later in vitamins and minerals. Fruits, vegetables, and animal-source foods (dairy, meat, fish and eggs) were considered "protective" against deficiencies (see Harper, 1987 for a history of dietary guidance).

However, as the food supply and food consumption patterns changed in many parts of the world, the second half of the 20th century brought new concerns and new diet quality concepts. Relationships between dietary patterns and non-communicable disease (NCD) risk were illuminated, and a large body of evidence emerged regarding a different type of "protective" diet, one protective against NCDs (Mozaffarian, Rosenberg, and Uauy, 2018)

Evidence of benefit accumulated for particular regional diets, which could be adapted to suit other settings (for example, the Mediterranean and "alternate" Mediterranean patterns), for diets designed to reduce hypertension (for example, the "Dietary Approaches to Stop Hypertension" (DASH) diet) and others focused on NCD risk reduction more broadly.

Other specific recent concerns related to diet quality include the inflammatory effects of dietary components and consequent NCD risk (Shivappa et al., 2014) and the impact of ultra-processed foods on nutrient intakes and health risks (Elizabeth et al., 2020). Food safety risks are not new but continue as a concern in all contexts. Beyond human health effects, there is increasing recognition of the effect of diets and food systems on planetary health, and these concerns are now incorporated in some definitions of diet quality (FAO & WHO, 2019).

Global concerns thus now include sufficiency/adequacy, NCD risk reduction, food safety, and sustainability, and these concerns are relevant in all countries (GBD 2017 Diet Collaborators, 2019).

8c.2.2 Basis for diet quality indices

Indices used for assessing diet quality may be based on national or global dietary guidelines, or they may be based on other assessments of evidence related to nutritional adequacy, diet-health relationships, and/or environmental sustainability. Depending on the underlying definition of diet quality as well as on intended uses, diet quality indices may be designed to capture one, several, or all of the following characteristics of diets:

- Food groups consumed: Quantities, diversity between and/or within food groups, and/or balance (proportionality) among food groups;
- Macronutrient balance (for example, percent of energy from protein, carbohydrates and fat);
- Nutrient density of the diet (for example, nutrients per 1000kcals);
- Nutrient adequacy of the diet relative to requirements;
- Moderation in consumption of foods and food groups that elevate NCD risk;
- Adherence to dietary patterns designed to minimize risk of NCDs;
- Adherence to defined healthy cultural dietary patterns;
- Environmental impact and/or sustainability of the diet;
- Adherence to dietary guidance from national or global health authorities, such as national FBDGs or World Health Organization (WHO) guidance.

There is much overlap among these characteristics. For example, measures of adherence to healthy cultural dietary patterns may capture food group diversity, proportionality, macronutrient balance, nutrient adequacy and NCD risk reduction. Examples include indices for regional diets such as Mediterranean diets (D'Alessandro & De Pergola, 2018) and "adapted" Mediterranean diets (Fung et al., 2006), Japanese diets (Kanauchi & Kanauchi, 2019), and Nordic diets (Hillesund et al., 2014).

As noted above, national FBDGs also generally have multiple objectives, including ensuring nutrient adequacy and reducing NCD risks. Increasingly, as noted in Section 8c.1, FBDGs may also aim to address environmental impacts of diets (Gonzalez Fischer & Garnett, 2016) and/or to specifically advise moderation in consumption of certain types of highly processed food (Ministry of Health of Brazil, 2014; Health Canada, 2019).

Thus, diet quality indices designed to reflect adherence to national or global guidelines will reflect the definitions of diet quality and the diverse aspects of diets and dietary patterns that are described in the particular guidelines.

8c.2.3 Uses of diet quality indices

Diet quality indices have been developed and used for a variety of purposes, including:

- Assessment and monitoring
  - Describing diet quality for an individual
  - Describing diet quality among a population
  - Estimating distributions of diet quality among a population
  - Describing trends in population diet quality over time
- Epidemiological research

    - Examining associations between diet quality and a dependent variable (for example, does adherence to a dietary pattern reduce risk of a particular NCD?)
    - Examining associations between an independent variable and diet quality (for example, do lower income groups have lower diet quality?)
- Programmatic and policy uses
    - Targeting programs
    - Assessing the effects of policy and programmatic interventions on diet quality
    - Advocacy and informing policy development

Just as with nutrient intake assessment (Chapter 8b), diet quality assessment often requires accounting for day-to-day variation in intakes. See Sections 3.3 and 8a.2.1 for discussions of appropriate dietary measurement and analytic methods for various objectives.

Certain characteristics of indices may be more or less important depending on intended uses. For example, when an index will be used to compare diet quality across countries (or across regions or cultures within countries), the index should have the same meaning across diverse cultural dietary patterns (Frongillo et al., 2019).

In some contexts, when an index is needed for frequent monitoring of trends (for example, annually or biannually), lower cost, lower burden indices may be required due to resource constraints. When an index will be used to summarize diet quality for the purposes of communicating or advocating with non-technical audiences (whether consumers or policymakers) the index should be amenable to simple, clear presentation and interpretation. Finally, for most uses, and particularly for use in assessing trends and in evaluating impacts of policy and programmatic interventions, indices should be responsive — that is, they should change in response to meaningful changes in diet quality.

8c.2.4 Development of diet quality indices

There are two general approaches to assessing dietary patterns of individuals, and to development of indices. One approach assesses the diet relative to a pre-defined set of criteria ("*a priori*" dietary patterns) such as national or global recommendations, or a defined cultural dietary pattern.

The second general approach ("*a posteriori*" dietary patterns, sometimes called "exploratory") is data driven, involving the examination of clustering of diet characteristics in a particular study sample (Gleason et al., 2015). There are several analytic approaches to defining a data-driven dietary pattern, including factor analysis, principal components analysis, and cluster analysis. Reduced rank regression is a hybrid approach, with a priori identification of characteristics combined with data-driven identification of patterns. Because data-driven dietary patterns are difficult to generalize beyond a particular study setting, they are not further discussed in this section.

For *a priori* dietary patterns, selection of index components (foods, nutrients, etc.) and quantitative criteria for intakes may be well-defined (for example, based on guidelines), but there are still many decisions to be taken during development of indices (Box 8c.4). In the next Section, many of these decisions will be illustrated with a detailed example describing development of the U.S. "Healthy Eating Index" (HEI), which measures adherence to the U.S. dietary guidelines.

Box 8c.4 Examples of decisions to be taken during development of diet quality indices

What are the component parts of the diet quality index?

- Should it include food items, food groups, nutrients, and/or food substances such as dietary fiber, polyphenols, or sugar?

When components are food groups, how are they defined? For example:

- Should potatoes or legumes "count" as vegetables?
- Should juice "count" as fruit?
- Should snails or insects "count" as "meat"?

Are energy intakes accounted for, to distinguish the impact of quantity of food from quality? If so, how? For example:

- Should food group intakes be “counted” in absolute terms, such as in grams? Or should food group intakes be counted in relative terms (for example, grams per 1000kcals of total intake)?

How are components scored? For example:

- Is scoring based on absolute criteria, or on relative criteria, such as study-specific distributions of intake for each component?
- Does scoring differ for different age/sex groups, or other groupings?
- How many points per component?
- For scoring based on criteria (not distributions), should scoring be:
  - Dichotomous (1 point if the criterion is achieved, and 0 points if not)?
  - Proportional (ranging from lowest score to highest depending on how closely intake adheres to the criterion or recommendation)?
  - Scored in some other way?

How are components weighted? How are sub-indices weighted? For example:

- Should all components be given the same “weight” in the total score? Or should some components be given heavier “weight” based on knowledge of associations to benefits or risks?
- For indices with both “adequacy” and “moderation” components, how should the total score for the set of all adequacy components be weighted relative to the total score for the set of moderation components?

Both during development of indices and afterwards, performance is evaluated in several ways. Indices should be valid and reliable and well-suited for intended uses. Validation of indices aims to answer the following questions, among others:

- Is the index well-grounded in theory?
- Does it cover all dimensions / characteristics of interest (content validity)?
- Does it perform in a way that is consistent with theory (construct validity)?

Kirkpatrick et al. (2019) provide a detailed discussion of validation issues for dietary assessment measures. For a discussion of validation concepts more broadly, see Frongillo et al. (2019).

This page titled 7.2: Assessing diet quality (8c.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.3: Examples of diet quality indices (8c.3)

This Section presents a series of specific examples of diet quality indices. Literature reviews have identified hundreds of diet quality indices; some are study- or context-specific, while others are proposed for national, regional or global use (Burggraf et al., 2018; Trijsburg et al., 2019; Aljuraiban et al., 2020; Miller et al., 2020). The purpose of this Section is to provide a small set of examples of diverse types of indices, with differing origins and intended uses.

Most indices have been developed for adult populations, though they have sometimes been adapted for children and adolescents. Indices developed or adapted for children and adolescents were recently reviewed by Dalwood et al. (2020).

Historically, most diet quality indices were initially developed in high-income countries and rely on quantitative dietary recall data from food frequency questionnaires and/or quantitative 24hr recalls for their measurement. These indices also require availability of detailed and comprehensive food composition data. However, several recent initiatives have aimed to develop lower burden proxy indices measuring one or more diet quality characteristics, with less costly data requirements, for use in global and/or national monitoring. Table 8c.1 presents some examples of some of the most widely used indices, as well as some newer indices intended for global use.

Table 8c.1 Examples of diet quality indices

Name	Basis of Index
Food and/or nutrient-based indices	
U.S. Healthy Eating Index (HEI)	Measures adherence to key recommendations in the Dietary Guidelines for Americans
Alternate Healthy Eating Index (AHEI)	Measures consumption of a literature-based selection of foods and nutrients consistently associated with decreased risk of NCDs
Mediterranean Diet Score (MDS)	Measures adherence to one of the traditional dietary patterns of the Mediterranean region, or adapted versions of these
Dietary Approaches to Stop Hypertension diet index (DASH)	Measures adherence to a dietary pattern originally developed in a randomized controlled trial for reducing hypertension
Healthy Diet Indicator (HDI)	Measures adherence to WHO global dietary recommendations for the prevention of chronic disease
Indices based on other characteristics	
Dietary Inflammatory Index (DII)	A scoring system summarizing the inflammatory potential of the diet; inflammatory diets have been linked to NCDs
Percent of energy from ultra-processed foods (% UPF)	Metric based on evidence that intakes of UPF are associated with poor quality diets and health risks
Lower-burden indices designed for settings where quantitative data are not feasible	
Minimum dietary diversity (MDD)	Proxy for micronutrient density of infant and young child diets
Minimum dietary diversity for women of reproductive age (MDD-W)	Proxy for micronutrient adequacy of women's diets
Global Dietary Recommendations Score (GDR)	Adherence to WHO Healthy Diet guidance and World Cancer Research Fund / American Institute for Cancer Research recommendations; proxy for HDI

Name	Basis of Index
Global Diet Quality Score (GDQS)	Measures consumption of a literature-based selection of food groups that contribute to nutrient intake and NCD risk reduction across a variety of global dietary patterns.

The first set of food and/or nutrient-based indices in Table 8c.1 — the HEI, AHEI, MDS, DASH, and HDI — have a longer history of use, and most have been updated several times to reflect evolutions in national or global recommendations and/or newer syntheses of epidemiological evidence. Figure 8c.2 shows results from simple searches of the PubMed database, showing the volume of studies mentioning each type of index before 2015 as compared to in the years 2015–2020. The U.S. Healthy Eating Index (HEI) has been widely used over several decades and is very well documented. Information on recommended presentation and interpretation is also available. Because of this, we use the HEI as a detailed example. Examples of other indices follow but are described in less detail.

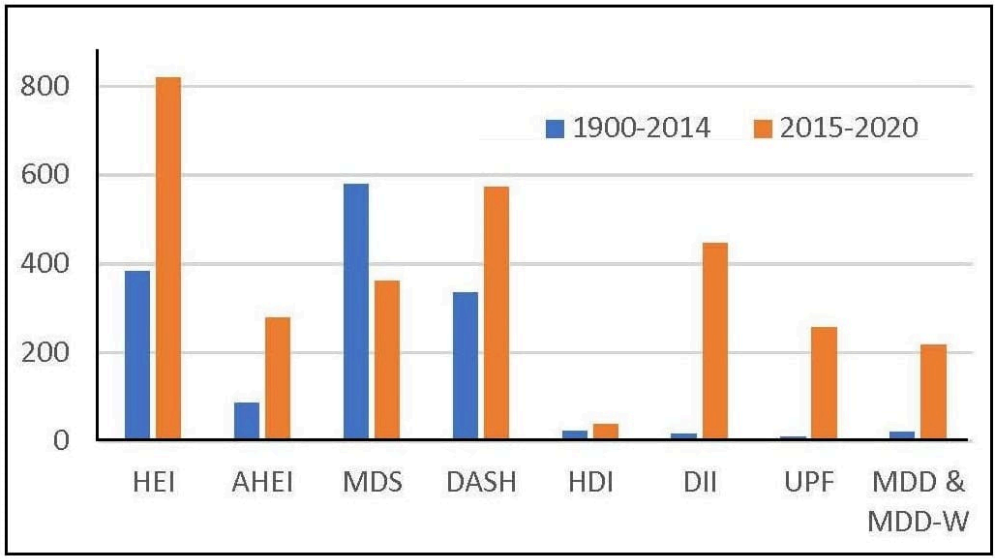


Figure 8c.2. Volume of studies mentioning various diet quality indices.

The PubMed database was searched on 15 February 2021, for the following terms, in all fields: for HEI, "Healthy Eating Index"; AHEI, "Alternate Healthy Eating Index"; MDS, "Mediterranean Diet Score"; DASH, "Dietary Approaches to Stop Hypertension"; HDI, "Healthy Diet Indicator"; DII, "Dietary Inflammatory Index"; UPF, "Ultra-processed Food"; MDD, "Minimum Dietary Diversity". Additional PubMed search terms were: for HEI: "NOT alternate NOT alternative"; for AHEI: "alternative healthy eating index"; for UPF: "NOVA classification". The GDR Score and the GDQS are not included because they are too recent; the GDR Score was published in November 2020 and the GDQS was published in 2021. Note that for the search related to the HEI, articles could have been reporting on the U.S. HEI, or on adaptations or similar indices developed for other countries but using the same index name.

8c.3.1 Basis, development, and scoring of the U.S. Healthy Eating Index

The Healthy Eating Index (HEI) was designed to reflect adherence to the U.S. national dietary guidelines, which apply to all individuals ≥ 2y. The dietary guidelines are intended both to ensure nutrient adequacy and to reduce risks of NCDs, and hence the HEI also reflects both adequacy and NCD risk reduction. The HEI can be used to characterize diets of children ≥ 2y as well as adults.

The US dietary guidelines are updated every five years. The HEI was first published in 1995 (Kennedy et al., 1995),and was updated in 2005, 2010 and 2015 to reflect new and revised guidance (Krebs-Smith et al., 2018); the HEI will be updated again in 2022.

The HEI can be calculated using data from quantitative 24hr recalls, food records, or food frequency questionnaires. The appropriate method of calculation depends on the data type and on intended uses. For most uses, the HEI should be calculated based on usual dietary intake (i.e., accounting for day-to-day variability in intakes) (Kirkpatrick et al., 2018).

The original HEI included 10 components: five food groups reflecting adequacy, four nutrients reflecting moderation, and a measure of variety in food intake. Each component was scored from 0 to 10, with adequacy components (food group scores) based on total intake, yielding a total possible score of 100. Table 8c.2 shows scoring for the newer HEI-2015, which includes nine adequacy components, and four moderation components.

Table 8c.2 Components and scoring standards for the Healthy Eating Index-2015. Source: U. S. Department of Agriculture Food and Nutrition Service

[a] *Intakes between the minimum and maximum standards are scored proportionately.*

[b] *Includes 100% fruit juice.*

[c] *Includes all forms except juice.*

[d] *Includes legumes (beans and peas).*

[e] *Includes all milk products, such as fluid milk, yogurt and cheese, and fortified soy beverages.*

[f] *Includes seafood, nuts, seeds, soy products (other than beverages), and legumes (beans and peas).*

[g] *Ratio of poly- and mono-unsaturated fatty acids (PUFAs and MUFAs) to saturated fatty acids (SFAs).*

Component	Max. points[a]	Standard for maximum score	Standard for minimum score
Adequacy			
Total fruits[b]	5	≥ 0.8 cup equivalent per 1000kcal	No fruit
Whole fruits[c]	5	≥ 0.4 cup equivalent per 1000kcal	No whole fruit
Total vegetables[d]	5	≥ 1.1 cup equivalent per 1000kcal	No Vegetables
Greens and beans[d]	5	≥ 0.2 cup equivalent per 1000kcal	No dark-green vegetable or legumes
Whole grains	10	≥ 1.5 ounce equivalent per 1000kcal	No whole grains
Dairy[e]	10	≥ 1.3 cup equivalent per 1000kcal	No dairy
Total protein foods[d]	5	≥ 2.5 ounce equivalent per 1000kcal	No protein foods
Seafood and plant proteins[d,f]	5	≥ 0.8 ounce equivalent per 1000kcal	No seafood or plant proteins
Fatty acids[g]	10	(PUFA + MUFA)/SFAs ≥ 2.5	(PUFA + MUFA)/SFAs ≤ 1.2
Moderation			
Refined grains	10	≤ 1.8 ounce equivalent per 1000kcal	≥ 4.3 ounce equivalent per 1000kcal
Sodium	10	≤ 1.1 grams per 1000kcal	≥ 2 grams per 1000kcal
Added sugars	10	≤ 6.5% of energy	≥ 26% of energy
Saturated fats	10	≤ 8% of energy	≥ 16% of energy

The scoring for the HEI illustrates some of the many decisions made during development of diet quality indices more generally. These include:

- Nature of component parts (for example, food groups, nutrients, or both);

- Operational definition of what food items “count” for each component.
- Decision on how to account for energy intakes;
- Criteria for scoring of each component, including quantity of intake for maximum score;
- Relative scoring of components (i.e., weighting), and range for total score;

Krebs-Smith et al. (2018) describe the rationale for all decisions taken during development of the HEI-2015. In brief, selection of component parts reflects key recommendations in the 2015–2020 Dietary Guidelines for Americans (U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2015). Footnotes to Table 8c.2 provide some information on how foods groups are defined and operationalized and additional details are available on the HEI website.

Except for the first (1995) version, HEI scoring has been based on amounts per 1000kcal (i.e., density), rather than on total amounts. One rationale for use of the density approach is that the HEI can then be applied at different levels (individuals, groups of people, food environments, food supply).

The density approach also has the effect of minimizing (though not eliminating) differences in food group quantity recommendations across differing energy intake levels (which in turn correspond to differences in age, size and activity level). To allow for consistent scoring, the HEI takes the least restrictive (easiest to achieve) target for density of intakes for energy intake levels ranging from 1200–2400kcal.

For most adequacy components (food groups), the maximum score is given for intake at or above the target density, and a minimum score of “0” is given for no intake. For fatty acid adequacy, and for the moderation components, the rationale for minimum and maximum scores is somewhat more complex (see Krebs-Smith et al., 2018 for details).

For all components, densities between the minimum and maximum standards are scored proportionally. For example, consuming no whole grains is scored at zero, and consuming 1.5 ounces of whole grains per 1000kcals is scored at the maximum of 10 points. Consuming 0.75 ounces per 1000kcals is scored at 5 points since it is halfway from zero to the target density.

While the HEI-2015 has more “adequacy” components than the original HEI (9 vs.5), the original total score of 100 is maintained. Certain original food groups are sub-divided: for example, the original HEI awarded up to 10 points for vegetables, whereas HEI-2015 awards 5 points for “greens and beans” and an additional 5 points for “total vegetables”.

Considering the five broader food groups (vs. subgroups, that is: fruits; vegetables; whole grains; dairy; and protein foods), each is equally weighted in HEI-2015, with a maximum score of 10 points. Fatty acids and each of the 4 moderation components also have maximum scores of 10 points. Overall, the adequacy components total 60 of the 100 points, whereas the moderation components total 40 of the 100 points.

The rationale for equal weighting of components is that the Dietary Guidelines for Americans are meant to be considered as a whole and that all concepts are equally important; further, there is currently no evidence to support unequal weighting (Krebs-Smith et al., 2018). Equal weighting of components is a feature of many other diet quality indices, and for the same reasons.

8c.3.2 Evaluation of U.S. Healthy Eating Index-2015

The validity of the HEI-2015 and its predecessors has has been assessed in several ways. Krebs-Smith et al. (2018) report on content validity — that is, the extent to which the index captures relevant dimensions of a healthy diet, as defined in the U.S. dietary guidance.

Reedy et al. (2018) report on construct and criterion-related validity. Construct validity was assessed in several ways. “Exemplary menus” were shown to receive high HEI-2015 scores; scores in a nationally representative sample demonstrated sufficient variation (range from the 1st–99th percentile was from 33 to 81 points) and scores were significantly different between smokers and non-smokers, as expected *a priori*. Scores were independent of energy intakes, i.e., reflective of quality, rather than quantity of intakes.

Criterion-related validity was demonstrated in a prospective cohort, where the highest vs the lowest quintiles of the HEI-2015 scores were associated with a decreased risk of all cause, cancer, and cardiovascular disease mortality. In addition, scores for previous versions of the HEI have been shown to be associated with reduced risks of NCDs and mortality in numerous studies (see, for example, Liese et al., 2015; Schwingshackl & Hoffmann, 2015).

8c.3.3 Interpretation of the U.S. Healthy Eating Index-2015

For descriptive purposes, total scores can be complemented by graphical representations and by a grading scale (Krebs-Smith et al., 2018). Radar graphs (Figure 8c.3) can be used to visualize population-level dietary patterns. In the radar graph, each component score is plotted as a percent of maximum points, on an axis. The perimeter illustrates a perfect score. Radar graphs are also useful because they show how the same total score can represent different patterns (sets of component scores).

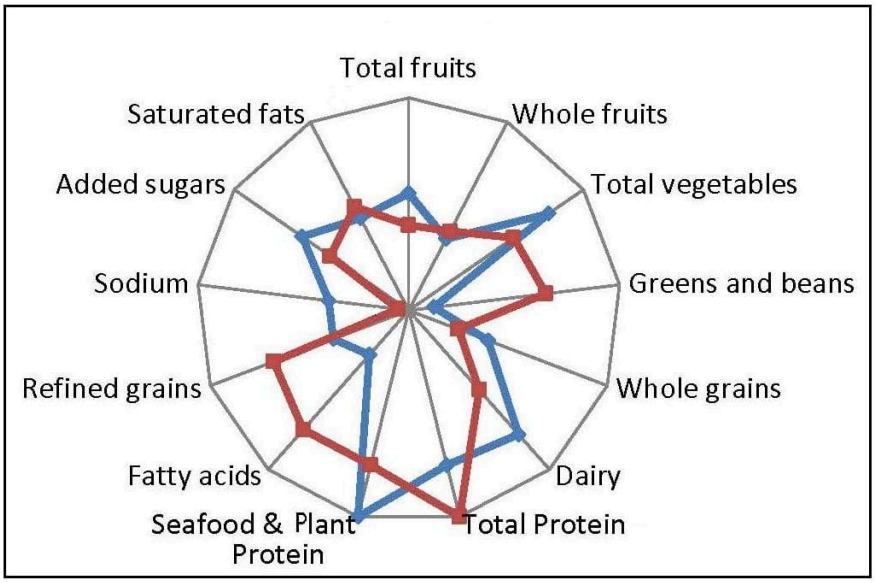


Figure 8c.3 Radar graph depicting a perfect score for the Healthy Eating Index 2015 (100 points) and two identical total scores (50 points) with different component scores. Redrawn from data in Krebs-Smith et al., 2018, p.1800.

To qualitatively describe population-level scores, a graded approach can be used, with 90 to 100 points an “A”, 80 to 89 a “B”, etc. Similarly, each component score can be graded (Krebs-Smith et al., 2018).However, categorized scores (translated into “grades”) should not be used for analysis purposes as categorization results in loss of information and may result in misclassification.

To interpret the magnitude of a difference in scores between groups, or across time in the same population, a difference (or change) of 5–6 points was shown to represent a “moderate” effect size for the U.S. population (Kirkpatrick et al., 2018).However, the authors caution that this may not apply to other populations, since effect sizes are calculated in relation to the typical dispersion of scores (standard deviations), which may vary across populations.

More generally, the radar graph or graded approach could be used for descriptive purposes with a variety of multi-component diet quality indices. Determining meaningful differences could similarly be informed by examination of standard deviations of the score in the population of interest, ideally in multiple studies. A difference (or change) in scores of approximately one-half of a standard deviation is considered to be a moderate effect size(Kirkpatrick et al., 2018).

Box 8c.5 summarizes key points about the HEI.

- Measures adherence to 2015–2020 U.S. dietary guidelines;
- Applies to children ≥ 2y, adolescents and adults;
- Component parts are food and beverage groups, nutrients (fatty acids, sodium), and added sugar;
- Components are expressed as densities (usually, per 1000kcal);

- Can be used to assess at various levels (for example, diet quality of populations, menus, food supplies);
- Requires detailed quantitative dietary intake data in grams or standard servings consumed, as well as linkages to databases with information on nutrients and food groups;
- Scoring is based on *a priori* criteria for intakes as found in the U.S. guidelines; scoring is not study-specific and relative (for example, quartiles);
- Intakes between minimum and maximum scores (per component) are scored proportionally;
- Sixty points are awarded for adequacy components and 40 points for moderation components;
- Guidance on presentation is provided, including visual presentation, use of “grades”, and the interpretation of effect sizes.

8c.3.4 Adaptation of the Healthy Eating Index concept

In addition to the periodic updates related to updated U.S. dietary guidelines, the HEI concept has also been adapted for other contexts. Similar indices measuring adherence to national guidelines were developed for many countries, including:

- **Australia (Roy et al., 2016)**
- **Brazil (Previdelli et al., 2011; Andrade et al., 2013)**
- **Canada (Garriguet, 2009; Jessri, Ng, and L'Abbé, 2017)**
- **China (Yuan et al., 2017; 2018)**
- **Denmark (Bjerregaard et al., 2019)**
- **France (Estaquio et al., 2009; Chaltiel et al., 2019)**
- **Germany (Moon et al., 2017; Mader et al., 2020)**
- **Korea (Yook et al., 2015; Park et al., 2020)**
- **Malaysia (Pondor, Gan, & Appannah, 2017)**
- **Mexico (López-Olmedo et al., 2019)**
- **The Netherlands (Looman et al., 2017)**
- **Russia (Martinchik et al., 2019)**
- **Spain (Norte Navarro & Ortiz Moncada, 2011)**
- **Thailand (Taechangam, Pinitchun, & Pachotikarn, 2008)**

In Ethiopia, development of food-based dietary guidelines is underway and an Ethiopian HEI will also be developed (Bekele et al., 2019). HEI were also created for children 1y and older in Finland (Kyttälä et al., 2014), pregnant women in Singapore (Han et al., 2015), and adolescents in Taiwan (Y. C. Chen et al., 2018). Many of these indices are also scored from 0 to 100, but some have different numbers of components and different ranges for scoring.

This page titled 7.3: Examples of diet quality indices (8c.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

While the HEI and similar indices in other countries measure adherence to current national dietary guidelines, another very widely used index, the "Alternate Healthy Eating Index" (AHEI), was developed to reflect additional evidence related to NCDs, which had not been incorporated in the US national guidance at the time.

As first developed by McCullough et al. (2002), the AHEI incorporated some of the same elements as the original HEI, but in addition emphasized: consumption of whole grains; a favorable ratio of fish and poultry to red and processed meats; intake of nuts and soy foods; and a favorable ratio of polyunsaturated to saturated fats. Unlike the HEI, dairy is not included in the scoring. The AHEI showed stronger associations than the original HEI to cardiovascular disease outcomes for both men (McCullough et al., 2000) and women (McCullough et al., 2000). The authors noted that this evidence could help inform future revisions of dietary guidelines.

Table 8c.3 Components and scoring standards for the Alternate Healthy Eating Index 2010. Adapted from Chiuve et al. (2012).
Several apparent errors in original paper were corrected in the table above, by referring to Liese et al. (2015), Supplemental material p. 6.
PUFA = polyunsaturated fatty acids.
[a] *Intakes between the minimum and maximum standards are scored proportionately.*
[b] *Includes all vegetables except potatoes. One serving is 0.5cup of vegetables or 1cup of green leafy vegetables.*
[c] *Does not include fruit juice. One serving is 1 medium piece of fruit or 0.5cup of berries.*
[d] *Use of grams of whole grains accounts for the variability of the percentages of whole grain in various. One serving of a 100% whole-grain product (i.e., 0.5cup of oatmeal or brown rice) contains ≈ 15–20g of whole grains (per dry weight).*
[e] *One serving is 8oz.*
[f] *One serving is 1oz of nuts or 1 tablespoon (15mL) of peanut butter.*
[g] *One serving is 4oz of unprocessed meat or 1.5oz of processed meat.*
[h] *Optimal intake (250mg/d) is ≈ two 4oz servings of fish per week.*
[i] *Cutoffs for sodium were based on deciles of distribution in the data sets used to develop the index. This was due to lack of brand specificity on the food frequency questionnaires, such that absolute intakes could not be accurately estimated. This implies users would also base cutoffs on deciles of distribution in the data set analyzed.*
[j] *Highest scores are for moderate alcohol intake, and the worst score is for heavy consumption. Nondrinkers receive an intermediate score of 2.5.*

Component	Standard for maximum. score (10)[a]	Standard for minimum score (0)
Vegetables, servings/d [b]	≥ 5	0
Fruit, servings/d [c]	≥ 4	0
Whole grains, g/d [d] — Women	75	0
Whole grains, g/d [d] — Men	90	0
Sugar-sweetened beverages and fruit juice, servings/d [e]	0	≥ 1
Nuts and legumes, servings/d [f]	≥ 1	0
Red/processed meat, servings /d [g]	≤ 1	≥ 1.5
Trans fat, % of energy	≤ 0.5	≥ 4
Long-chain (n-3) fats (EPA + DHA), mg/d [h]	250	0
PUFA, % of energy	≥ 10	≤ 2
Sodium, mg/d [i]	Lowest decile	Highest decile
Alcohol, drinks/d [j] — Women	0.5–1.5	≥ 2.5
Alcohol, drinks/d [j] — Men	0.5–2.0	≥ 3.5

Component	Standard for maximum. score (10)[a]	Standard for minimum score (0)
Total points	110	0

Like the HEI, the AHEI was later revised to reflect new epidemiological evidence on relationships of dietary components to NCD risks, and the newer version is referred to as the AHEI 2010 (Chiuve et al., 2012). As in the original version, the rationale for the AHEI 2010 was a comprehensive review of clinical and epidemiological research by the authors, with selection and scoring of components that were deemed to be consistently associated with NCD risk. As with the original, the AHEI 2010 did not fully align with the HEI.

Scoring for the AHEI 2010 is shown in Table 8c.3. Both versions of the AHEI were developed using semi-quantitative food frequency data from prospective cohort studies of U.S. health professionals. However, subsequent users have also calculated scores for the AHEI from quantitative 24hr recall data and for other population groups, including adolescents (see, for example Dalwood et al., 2020; Ducharme-Smith et al., 2021); however, Dalwood et al. caution against application of indices developed for adults in pediatric populations.

Like the HEI, the AHEI includes both foods and nutrients (fatty acids) among the component parts, but it also includes sweet beverages and alcohol. Unlike the HEI, the AHEI is not scored based on densities (quantities per 1000kcal), but instead based on total consumption per day. Unlike in the HEI, fruit juice is included with sugar-sweetened beverages rather than with fruit.

Highest scores are based on literature-based judgments of minimum risk. Consistent with the HEI, intakes that are intermediate between the minimum and maximum are generally scored proportionally. However, alcohol has "U-shaped" scoring, such that the highest score is given for moderate intake and lower scores for "0" or for excessive intake. Sodium intakes are scored based on the study specific distribution. Also as for the HEI, components each have a maximum of 10 points, in this case yielding a range of up to 110 points (11 components), with 6 "positive" components, 4 "negative" components and alcohol (U-shaped). Box 8c.6 summarizes key points about the AHEI.

Box 8x.6 Summary of Alternate Healthy Eating Index 2015

- Measures adherence to a set of recommendations based on review of epidemiological evidence by researchers;
- Developed for adults;
- Component parts are food and beverage groups and nutrients (fatty acids), mostly expressed as quantities consumed per day;
- Requires detailed quantitative dietary intake data (grams or standard servings consumed); as well as linkages to databases with information on nutrients and food groups;
- Scoring is based on *a priori* criteria for intakes with the exception of sodium, which is scored based on deciles in the data set under study;
- Intakes between minimum and maximum scores (per component) are scored proportionally;
- Sixty points are awarded for adequacy components and 40 points for moderation components; for alcohol, scoring is "U-shaped"

This page titled 7.4: Alternate Healthy Eating Index (8c.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

Evidence on the healthfulness of various Mediterranean regional diets has been accumulating since the mid-20th century (Menotti & Puddu, 2015).Generally, Mediterranean diets are characterized by a high intake of plant-based foods including olive oil, fruit, nuts, vegetables, and cereals; a moderate intake of fish and poultry; a low intake of dairy products, red meat, processed meats, and sweets; and a moderate amount of wine. However, unlike the dietary patterns captured by the HEI and AHEI, which each have a single definition, there are numerous Mediterranean diets and numerous scores and indices (see: Milà-Villarroel et al., 2011; D'Alessandro & Pergola, 2018; Zaragoza-Martí et al., 2018).

In contrast to the HEI and AHEI, Mediterranean Diet Scores (MDSs) are generally based on foods and beverages only, rather than on a mix of foods/beverages and nutrients. Intended uses of MDSs are not consistently clearly stated in papers reporting on their development. They are frequently used in epidemiological studies investigating associations of diets to morbidity and mortality. MDSs vary widely in their scoring range (for example, possible scores of 0–8 vs. 0–100) and in the rationale and method of scoring, as illustrated in the following examples.

8c.5.1 Mediterranean Diet Score of Trichopoulou

MDSs with the narrowest ranges are scored as “yes” (1) or “no” (0) for a small set of food groups. Trichopoulou et al. (1995) developed one of the first MDS and demonstrated an association with mortality in a Greek population. The original analysis was based on semi-quantitative food frequency data from adults over 70y. This MDS was designed for an epidemiological study. The MDS of Trichopoulou et al. is an 8-point score, assessing for the following:

- High consumption of legumes
- High consumption of cereals (including bread and potatoes)
- High consumption of fruits
- High consumption of vegetables
- Low consumption of meat and meat products
- Low consumption of milk and dairy products
- High ratio of monounsaturated to saturated fat
- Moderate alcohol consumption

This score has 5 “positive” 2 “negative” and one “U-shaped” component like the AHEI, for moderate alcohol intake. Dairy is included as a “negative” component; this is typical of many MDSs. “High” intake was defined as intake above the sex-specific median quantity consumed by their study subjects rather than in relation to a set of recommended intakes, and therefore scoring for this MDS is dependent on study-specific distributions of intake. This approach to scoring is useful for study-specific comparisons but less useful for comparing across contexts. Several other MDSs are scored in this way.

8c.5.2 “MedDietScore” of Panagiotakos

For a prospective cohort study of adults in Athens, Panagiotakos et al. (2007) developed a 55-point “MedDietScore” where scoring did not depend on the distribution of intake among study subjects. Instead, each component was scored from 0–5 based on respondent report of frequency of consumption, based on data from a semi-quantitative food frequency questionnaire covering the past year.

Component scores increased with frequency for 7 “positive” components — non-refined cereals, fruits, vegetables, legumes, olive oil, fish and potatoes — and decreased with frequency for 3 “negative” components — red meat and products, poultry and full-fat dairy products. As with several other indices, there were more “positive” components than “negative” components.

This MedDietScore was negatively associated with hypertension, hypercholesterolemia, diabetes and obesity in the initial study and was later shown to be negatively associated with 10-year incidence of cardiovascular disease (Panagiotakos et al., 2015) and diabetes (Filippatos et al., 2016).

8c.5.3 "MEDI-LITE" Score of Sofi

Sofi et al. (2014) developed an 18-point "literature-based" adherence score, shown in Table 8c.4. This adherence score, the "MEDI-LITE" score, was designed for both epidemiological and clinical use; however, a primary aim was to develop a simple tool feasible for use at the individual level, within clinical practice. Component scoring was based on a set of prospective cohort studies, published between 2010 and 2013. The studies were primarily from Europe and included the multi-country European Prospective Investigation into Cancer and Nutrition (EPIC) study as well as additional studies from some of the EPIC study countries (Denmark, France, Germany, Greece, Italy, The Netherlands, Norway, Spain, Sweden, and the UK). Several studies were from the U.S., and no studies were from Asia, Africa, or Latin America. All studies included adult participants only.

Table 8c.4 Components and scoring standards for MEDI-LITE. Adapted from Sofi et al. (2014),
Note these authors use the term "portion" in the way the term "serving" is used here — that is, as a defined standard quantity.

Component	Low intake (score)	Moderate intake (score)	High intake (score)
Fruit (1 portion = 150g)	< 1 portion/d (0)	1–1.5 portions/d (1)	> 2 portions/d (2)
Vegetables (1 portion = 100g)	< 1 portion/d (0)	1–2.5 portions/d (1)	> 2.5 portions/d (2)
Legumes (1 portion = 70g)	< 1 portion/wk (0)	1–2 portions/wk (1)	> 2 portions/wk (2)
Cereals (1 portion = 130g)	< 1 portion/d (0)	1–1.5 portions/d (1)	> 1.5 portions/d (2)
Fish (1 portion = 100g)	< 1 portion/wk (0)	1–2.5 portions/wk (1)	> 2.5 portions/wk (2)
Meat/meat products (1 portion = 80g)	< 1 portion/d (2)	1–1.5 portions/d (1)	> 1.5 portions/d (0)
Dairy products (1 portion = 180g)	< 1 portion/d (2)	1–1.5 portions/d (1)	> 1.5 portions/d (0)
Alcohol (1 alcohol unit (AU) = 12g)	< 1 AU/d (1)	1–2 AU/d (2)	> 2 AU/d (0)
Olive oil	Occasional use (0)	Frequent use (1)	Regular use (2)

For each component, intakes were categorized into three levels after calculating weighted averages (weighted by study sample size) and standard deviations of intakes across all studies. The score includes six "positive" components, two "negative" components, and U-shaped scoring for alcohol.

Construct validity for the MEDI-LITE score was demonstrated by comparing it to the previously validated MedDietScore of Panagiotakos (above).

8c.5.4 Mediterranean-Style Dietary Pattern Score of Rumawas

More recently, some Mediterranean diet indices have been proposed with wider scoring ranges and with scoring based on criterion values (recommended intakes). Rumawas et al. (2009) developed a 100-point Mediterranean-Style Dietary Pattern Score (MSDPS) based on the Mediterranean Diet Pyramid of Willett et al. (1995); scoring is shown in Table 8c.5. The MSDPS was explicitly intended for use in non-Mediterranean populations. During development, performance of the MSDPS was assessed among adults using data from a semi-quantitative food frequency questionnaire covering the past year.

Table 8c.5 Components and scoring standards for the Mediterranean-Style Dietary Pattern Score. Adapted from Rumawas et al. (2009), p. 1152.

[a] *Each component of the score is calculated based on the recommended intakes of food in the Mediterranean Diet Pyramid of Willett et al. (1995).*

[b] *Except olive oil, all other components are continuous, ranging from 0–10 and computed proportionately.*

If consumption exceeds the recommended intake, the score is deducted proportionally to the number of servings consumed that exceed the recommended intake; the lowest possible score due to deduction is zero.

Component	Standard for maximum score of 10[a]	Score[b]

Component	Standard for maximum score of 10[a]	Score[b]
	Servings/day	Points/serving
Whole grains	8	1.25
Fruits	3	3.33
Vegetables	6	1.67
Dairy	2	5.0
Wine - Women	1.5	6.67
Wine - Men	3	3.33
	Servings/week	Points/serving
Fish and other seafood	6	1.67
Poultry	4	2.5
Olives, legumes, and nuts	4	2.5
Potatoes, starchy roots	3	3.33
Eggs	3	3.33
Sweets	3	3.33
Meat	1	10.0
	Practice	Points
Olive oil	Use only olive oil	0 (for no use) 5 (for use of olive and other vegetables oils)

The thirteen components are each scored from 0–10, then summed. The total score is standardized to 100 by dividing the calculated sum by the theoretical maximum sum of 130 and multiplying by 100. To account for foods not on the Mediterranean Diet Pyramid, the standardized score is weighted by the proportion of total energy intake derived from foods on the pyramid. For example, if 20% of kilocalories are from foods not on the pyramid, the standardized score is multiplied by 0.8.

As for the HEI and AHEI, intakes that are lower than recommended are scored proportionately to the recommended amounts. Scoring is for total quantities (as for the AHEI) and not for densities (as for the HEI).

Unlike some other indices, scoring for the MSDPS "penalizes" overconsumption of a given food group by subtracting points in proportion to the amount of excess — for example, exceeding the recommended intake by 60% would result in a score of 4 for a given component. The lowest possible score for each component is zero. This means, for example, that both 2 servings of meat/week and 10 servings of meat/week are scored identically as zero.

Construct validity was demonstrated through expected positive associations with intakes of dietary fiber, fatty acids, antioxidant vitamins, calcium, magnesium, and potassium, and inverse associations with a glycemic index and with intakes of added sugar, saturated fat, and *trans*-fat.

8c.5.5 Summary of Mediterranean Diet Scores

There is no consensus on a single or "best" Mediterranean Diet Score or Index, and this depends on intended uses; see D'Alessandro and De Pergola (2018) for a discussion. Although many MDSs were originally developed based on analyses of data from adult populations, including elders, MDSs have been adapted and used in studies of children and adolescents (Dalwood et al., 2020).

Finally, concerning the potential to adapt the MDS for relevance in other regions, there is more experience in high-income countries and in the global north compared to the global south. The applicability of the pattern described by the MDS to some other regions is not established, particularly in regions where some of the distinctive components (for example, olive oil, red wine, nuts, legumes, and/or fish) are not available and/or are not part of cultural dietary patterns.

However, reporting on a global World Heart Federation Consensus Conference, Anand et al. (2015) suggest that the evidence base for the Mediterranean diet is sufficiently consolidated that it can be recommended globally, and they provide suggestions for its adaptation with foods from other regions. Once the dietary pattern is adapted to other regions, adapted scores could also be developed (as, for example, in Echeverría et al., 2019; El Kinany et al., 2020). Box 8c.7 summarizes key points about MDSs.

Box 8c.7 Summary of Mediterranean diet scores and indices

- There are a large number of scores and indices measuring adherence to different versions of the Mediterranean diet;
- The original scores were developed based on data from older adults, but there are many versions adapted for children and adolescents;
- MDSs have been adapted for several other regions, but certain characteristic foods are not consumed or are rarely consumed in some other global regions;
- Component parts are usually food groups; olive oil and alcohol are also often included;
- Components are usually expressed as quantities, not densities;
- Most MDSs requires detailed quantitative dietary intake data (grams or standard servings consumed);
- Most MDSs were developed for use in epidemiological studies, but are also appropriate to use in describing diet quality of populations;
- Scoring for components is sometimes absolute (based on *a priori* criteria for intakes) but for many MDSs, scoring is not absolute and instead is based on distributions of intake in the study sample (for example, with high points given to the highest quartile or quintile of intakes);
- Both the complexity of scoring and the balance between positive and negative scores vary widely among MDSs.

This page titled 7.5: Mediterranean Diet Score or Index (8c.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.6: Dietary Approaches to Stop Hypertension scores

Scores reflecting adherence to the Dietary Approaches to Stop Hypertension (DASH) dietary pattern have also been widely used, particularly in studies investigating associations between dietary patterns and NCDs. The original DASH diet and scoring were developed for a randomized controlled feeding trial investigating the impact of a prescribed dietary pattern on blood pressure. The trial, initiated in 1994, compared a typical U.S. dietary pattern to one high in fruits, vegetables, and low-fat dairy products, that emphasized fish and chicken rather than red meat, and was low in saturated fats, cholesterol, sugars, and refined carbohydrate.

The original study demonstrated reductions in blood pressure after the 8-week intervention (Sacks et al., 1999). Follow-up research demonstrated additional benefits from reductions in sodium intake; sodium intake had not been manipulated in the original trial (Sacks et al., 2001). While the DASH diet was originally developed to lower blood pressure, subsequent studies have linked higher adherence to the DASH diet with lower risks for a range of adverse outcomes, including all-cause mortality, cardiovascular disease incidence or mortality, cancer incidence or mortality, type 2 diabetes, and neurodegenerative disease (Morze et al., 2020).

As for Mediterranean diets, there have been several specific scores developed to reflect adherence to the DASH diet, and DASH scores have also been developed for pediatric populations. Table 8c.6 shows one commonly used scoring system with 8 components, each worth 5 points, for a total of 40 points. The scoring system is based on sex-specific quintile rankings within the study population.

Table 8c.6 Components and scoring standards for an example Dietary Approaches to Stop Hypertension diet adherence score. Scoring as in Fung et al., 2008 andLiese et al., 2015.

Component	Standard for maximum score	Standard for minimum score
Whole grains	Highest quintile	Lowest quintile
Vegetables (excluding potatoes)		
Fruit		
Nuts and legumes		
Low-fat dairy		
Red and processed meat	Lowest quintile	Highest quintile
Sugar-sweetened beverages		
Sodium		

With the exception of sodium, this DASH adherence score is based on foods rather than nutrients. As with the HEI, AHEI and MDS above, the eight components are given equal weight. As with many of the score described above, there are slightly more “positive” than “negative” components, implying an overall heavier weighting of positive food groups. Unlike in many MDSs, low-fat dairy is scored positively.

Like some versions of the MDS, and also like some other versions of DASH diet scores, scoring is dependent on study-specific distributions of intake. This type of scoring can maximize contrasts for studies investigating associations between dietary patterns and outcomes. However as noted above, it is less useful for comparing intakes across contexts.

This page titled 7.6: Dietary Approaches to Stop Hypertension scores is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.7: Healthy Diet Indicator (8c.7)

Just as the HEI was developed to capture adherence to U.S. national dietary guidance, the Healthy Diet Indicator (HDI) was developed to capture adherence to global guidance from the WHO (Huijbregts et al., 1997) and was revised to reflect updated guidance (Stefler et al., 2014; Jankovic et al., 2015). Both the original and the updated versions demonstrated associations with mortality in multi-country studies (Huijbregts et al., 1997; Stefler et al., 2014), although the associations were not entirely consistent (Jankovic et al., 2015).

The original HDI study analyzed data from cohorts of older men in three western European countries (Huijbregts et al., 1997), while the revised version was first used in studies of adults from Eastern Europe (Stefler et al., 2014) and elders from the U.S. and Europe (Jankovic et al., 2015). To our knowledge, the HDI has not been widely used in studies of children.

Unlike all the previous scores and indices, most of the components in the HDI are nutrients or food substances rather than food groups. This reflects the fact that the WHO "Healthy Diet Fact Sheet" currently gives quantitative guidance for intakes of sugars, various types of fatty acids, and salt, but limited guidance on quantitative intakes of food groups. The original HDI included 9 components, each scored "0" for not meeting or "1" for meeting the WHO recommendation, for a total range of 0–9 points. Components included: saturated fatty acids; polyunsaturated fatty acids; protein; "complex carbohydrates"; and mono- and disaccharides (all as percent of energy), and quantities for cholesterol; fiber; fruits and vegetables; and pulses / nuts / seeds. As with the indices described above, the components were equally weighted. The updated HDI included similar components, but each component was scored from 0 to 10, to better capture variation among study subjects.

Table 8c.7 Components and scoring standards for the Healthy Diet Indicator 2020. Eleven components are scored "0" or "1". Components 1–5 are used in a sub-index for healthy dietary components (maximum score, 5), and components 6–11 are used in a sub-index for dietary components to limit (maximum score, 6).

[a] *Upper end of the recommendation to consume no more than 350–500 grams/week divided by 7 days. Adapted from Herforth et al (2020).*

	Component	Standard for scoring (quantitative intake in one day)
1	Fruits, vegetables	≥ 400g
2	Beans and other legumes	> 0g
3	Nuts and seeds	> 0g
4	Whole grains	> 0g
5	Dietary fiber	> 25g
6	Total fat	< 30% total energy
7	Saturated fat	< 10% total energy
8	Dietary sodium	< 2g
9	Free sugars	< 10% total energy
10	Processed meat	0g
11	Unprocessed red meat	≤ 71g[a]

Most recently, Herforth et al. (2020) have again updated the HDI, reflecting the current WHO Healthy Diet recommendations and incorporating World Cancer Research Fund / American Institute for Cancer Research (WCRF / AICR) recommendations on the consumption of red and processed meats. Table 8c.7 shows the components and scoring for the HDI-2020.

Eleven components are scored "0" or "1", for a scoring range of 0–11. Several positive components are scored "1" for any non-zero consumption, based on WHO guidance that indicates that healthy diets include these items, but without specifying recommended

quantities (whole grains, beans and other legumes, and nuts and seeds). Components are equally weighted. Unlike many of the examples above, the total weight for all moderation / unhealthy components is slightly higher than the weight for healthy components. The revised HDI-2020 has not yet been evaluated for associations with health outcomes. Box 8c.8 summarizes key points about HDIs.

Box 8c.8 Summary of several versions of a Healthy Diet Indicator

- All versions of the HDI measure adherence to World Health Organization guidance; the most recent version also incorporates guidance from the World Cancer Research Fund/American Institute for Cancer Research;
- Focus is on NCD risk;
- Based on guidance for adults;
- Component parts are food groups, fiber, nutrients (fatty acids, sodium) and free sugars;
- Components are expressed in absolute terms, not as densities;
- Requires detailed quantitative dietary intake data (grams or standard servings consumed); as well as linkage to databases with information on nutrients
- Scoring is based on *a priori* criteria for intakes as found in global guidance; scoring is not study-specific and relative (for example, quartiles)
- In the HDI-2020, there are five healthy components and six components to limit, all weighted equally;
- In the HDI-2020, scoring is a simple count — with each of eleven components scored "0" or "1"

This page titled 7.7: Healthy Diet Indicator (8c.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.8: Dietary Inflammatory Indices (8c.8)

Recently, several diet quality indices have been developed to capture the inflammatory potential of the diet (Tabung et al., 2016; 2017; Hébert et al., 2019). Such indices were developed in recognition of the key role of chronic low-grade inflammation in the etiology of NCDs, and the role of dietary factors in chronic inflammation. In this section, we describe one widely used index, the Dietary Inflammatory Index (DII).

The DII was designed for global relevance, regardless of the cultural dietary pattern (Hébert et al., 2019). The DII is scored primarily based on nutrients and food substances. Selection of and weighting of components were based on a comprehensive literature review spanning 1950–2010 and including studies from diverse populations that assessed the relationships of nutrients, foods, and food substances to six inflammatory biomarkers (Cavicchia et al., 2009; Shivappa et al., 2014).

The DII has been applied in studies of a wide range of NCDs and risk factors. A recent review documented associations between DII scores and certain cancers, cardiovascular disease and its associated mortality, adverse mental health, and musculoskeletal disorders. Evidence on DII and respiratory health, neurodevelopmental outcomes, the metabolic syndrome, obesity and diabetes is either conflicting or limited to date (Phillips et al., 2019).

Scoring for the DII is based on 45 components, consisting mainly of macro- and micronutrients and food substances (for example, flavonoids, caffeine) but also including several foods, herbs/spices (garlic, ginger, onion, pepper, rosemary, saffron, thyme, and turmeric), and beverages (tea and alcohol). In the most recent version, scoring cut-offs were determined based on associations of components to inflammation in studies from eleven countries in all global regions except Africa. (Shivappa et al., 2014).

Each of the 45 components has a score indicating its pro- or anti-inflammatory potential, with positive scores for pro-inflammatory components and negative scores for anti-inflammatory components. The DII is the sum of the component scores. Unlike many indices described above, the components of the DII are not equally weighted. The total range in seven “example scenarios” was from +7.98 (strongly pro-inflammatory) to −8.87 (i.e. strongly anti-inflammatory); the example scenarios were created by the authors to illustrate various combinations of plausible intakes for the 45 components (Shivappa et al., 2014). Note that — unlike other indices — a negative score is better than a positive one.

The DII has been further adapted in a version that adjusts for total energy (E-DII) and in a version for children (C-DII), validated for children 6–14y. The C-DII predicted blood concentration of one inflammatory marker (C-reactive protein) in children (Khan et al., 2018). Construction of the DII requires data on all components, though it has been adapted for use when data for some components are not available (Davis et al., 2021).

The DII has primarily been used to further elucidate the relationship between the inflammatory potential of overall diets and various health outcomes. Box 8c.9 summarizes key points about the DII.

Box 8c.9 Summary of the Dietary Inflammatory Index

- The DII measures the inflammatory potential of the diet;
- Originally developed for adults, but a version has been developed for children 6–14y
- Component parts are primarily nutrients, food substances, and herbs and spices, with few foods or beverages — and no food groups — in the index;
- The DII requires detailed quantitative dietary intake data, including on intake of herbs and spices, and the availability of food composition data that includes many food substances;
- To date, the primary use of the DII has been in epidemiological studies.

This page titled 7.8: Dietary Inflammatory Indices (8c.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.9: Diet quality based on level of food processing (8c.9)

Another recent approach to assessment of diet quality is based on classification of foods and beverages according to the level of processing in their production. Monteiro et al. (2010) first proposed a system of classification, which was subsequently modified to include four groups:

- Unprocessed or minimally processed foods;
- Processed culinary ingredients;
- Processed foods; and
- Ultra-processed food and drink products.

The modified classification scheme is referred to as the "NOVA" classification (Monteiro et al., 2016). Particular attention has been paid to the role of ultra-processed products (UPF) in diets. While the classification is at the level of food and beverage items, the quality of the overall diet is then characterized by the proportion of UPF in the diet (usually, as a percent of total energy) (Monteiro et al., 2011; Chen et al., 2020).

Machado et al. (2019) describe UPF as "formulations of low-cost ingredients, many of non-culinary use, that result from a sequence of industrial processes (hence ultra-processed)". The identification and classification of UPF have recently been clarified:

> "A practical way to identify an ultra-processed product is to check to see if its list of ingredients contains at least one item characteristic of the NOVA ultra-processed food group, which is to say, either food substances never or rarely used in kitchens (such as high-fructose corn syrup, hydrogenated or interesterified oils, and hydrolysed proteins), or classes of additives designed to make the final product palatable or more appealing (such as flavours, flavour enhancers, colours, emulsifiers, emulsifying salts, sweeteners, thickeners, and anti-foaming, bulking, carbonating, foaming, gelling and glazing agents)" (Monteiro et al., 2019, p. 936).

Examples of UPF include packaged instant soups and noodles, carbonated beverages, reconstituted meat products, many sweet and savory packaged snacks, frozen "ready meals", and fast food dishes. Some UPF provide substantial nutrients, either from food ingredients (for example, packaged sweetened yogurts) or from fortification (for example, highly processed but fortified breakfast cereals), leading to challenges on definition of categories, and some disagreement on the value of the NOVA classification (Gibney, 2019; Jones, 2019).

The percent of energy from UPF has been shown to be high or very high in some countries; for example, UPF comprise from 40%–50% of energy intake in Australia (Machado et al., 2019), Brazil (Siqueira et al., 2020) and Canada (Polsky, Moubarac, & Garriguet, 2020) and nearly 60% in the United States (Martínez-Steele et al., 2017). While similar intake data are not available for most low- and middle-income countries, sales of UPF are growing rapidly in Asia and Africa (Vandevijvere et al., 2019; Baker et al., 2020). Assessment of the role of UPFs in the diet has been undertaken for a wide range of age groups, beginning in infancy.

A higher proportion of total intake from UPF has been associated with lower intakes of some micronutrients and higher intakes of sugars, salt, and saturated and trans-fatty acids (see for example Martínez-Steele et al., 2017; Moubarac et al., 2017; Machado et al., 2019). However, several studies documented higher calcium intakes with higher UPF consumption, perhaps due to consumption of sweetened and/or other processed dairy products(Louzada et al., 2015; Batal et al., 2018; Cornwell et al., 2018). Recent reviews have summarized associations with a wide range of risk factors and negative health outcomes, including overweight, obesity and cardiometabolic risks, some cancers, type 2 diabetes and cardiovascular diseases, irritable bowel syndrome, depression and all-cause mortality (Chen et al., 2020; Elizabeth et al., 2020). Box 8c.10 summarizes key points about the percent of energy from UPF.

Box 8c.10 Summary of percent of energy from ultra-processed foods

- Differs from other indices through its focus on one characteristic, processing level and this characteristic cuts across foods, beverages and food/beverage groups;
- Classification by processing level is done at the level of individual food and beverage item;
- Does not involve scoring based on consumption of food groups, nutrients, or food substances;

- The percent UPF is a density indicator, because it is expressed as a percent of energy;
- Like the HEI, it can be used for different age groups and at different levels, to characterize diets, menus, food supplies etc.;
- Requires detailed quantitative dietary intake data (grams or standard servings consumed).

This page titled 7.9: Diet quality based on level of food processing (8c.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.10: Lower-burden diet quality indices (8c.10)

Most diet quality indices described above require quantitative dietary intake data, for example from food frequency questionnaires, quantitative 24hr recalls, or other detailed quantitative approaches. Obtaining accurate quantitative dietary intake data is resource intensive. Because of the resource requirements, many low- and middle-income countries do not currently have nationally representative dietary intake data, and even high-income countries may not collect such data sufficiently frequently to meet all needs, such as for monitoring of population-level diet quality trends.

To meet needs for basic information on intake patterns in contexts where quantitative data are not available, several "lower-burden" approaches have been developed, particularly for use in global and national monitoring under resource constraints. These lower-burden approaches are all food group- based and/or ask behavioral questions (rather than including nutrient intakes for their calculation) and do not require food composition data for tabulation.

8c.10.1 Global surveillance of dietary behaviors

Many lower-burden survey tools (often called "screeners") have been developed to capture only one or several components of intake, and/or dietary habits. Examples of screeners developed for global use are the questionnaires for the WHO Stepwise Approach to NCD Surveillance (STEPS) and the Global School-Based Student Health Survey (GSHS). Currently, the WHO STEPS questionnaire includes questions on fruit, vegetable, and salt/salty condiment consumption. The GSHS questionnaire is under revision; previously, it included core questions on fruits, vegetables, carbonated soft drinks, and food from fast food restaurants. To date, neither of these surveys captures the "whole of diet", and data from these surveys have not been used to develop overall indices of diet quality.

8c.10.2 Food group diversity proxy indicators for micronutrients

Some food group diversity indicators have demonstrated associations with micronutrient adequacy of diets (Verger et al., 2021). Because of this, several have been developed as simple proxies for micronutrient adequacy, for use when more complex indices are not feasible. These indicators do not require use of food composition data, and they do not require detailed quantitative dietary intake data.

They were developed primarily for use in resource-constrained environments, where diets may be very impoverished, low in food group diversity, and consequently provide inadequate micronutrients. They are intended for population-level use — for example, in monitoring trends — and not to describe diets of individuals.

Indices were developed for several demographic groups, including infants and young children and women of reproductive age. In both cases, construct validity was evaluated cross-sectionally by comparing the simple indices to micronutrient density or adequacy of diets, in data sets from studies in multiple low- and lower-middle income countries. (Working Group on Infant and Young Child Feeding Indicators, 2006; Arimond et al., 2010;Martin-Prevel et al., 2017).

A "Minimum Diet Diversity" (MDD) indicator was developed as a proxy indicator of the micronutrient density of the diet for infants and young children 6–23mos (WHO 2021). MDD data can be collected using a lower-burden non-quantitative recall questionnaire for food groups fed to the infant or young child the day before the survey.

The updated version of the 2008 document includes 8 food groups. Infants and young children fed five or more of the eight food groups, in any amount, meet the criterion for MDD. Groups of infants and young children meeting MDD are likely to have a diet higher in micronutrient density than those who consume fewer food groups. Food groups are:

- Breast milk
- Grains, white/pale starchy roots, tubers and plantains
- Beans, peas, lentils, nuts and seeds
- Dairy products (milk, infant formula, yogurt, cheese)
- Flesh foods (meat, fish, poultry, organ meats)
- Eggs
- Vitamin A-rich fruits and vegetables

- Other fruits and vegetables

The MDD has been widely used and reported, with the main global data sources being the Demographic and Health Surveys (DHS) Program and the UNICEF Multiple Indicators Cluster Surveys (MICS). Prevalence data for MDD are available at the UNICEF infant and young child feeding indicators database.

Recently, the World Health Organization and UNICEF have developed additional feeding indicators relevant to diet quality of infants and young children, including indicators of sweet beverage consumption and consumption of selected unhealthy food groups that may displace nutrient-dense foods in their diets. However, there is no overall summary index for diet quality that covers both micronutrient density and non-recommended food groups (WHO/UNICEF, 2021).

A simple food group diversity indicator, MDD-W, was also developed for non-pregnant women of reproductive age as a proxy for micronutrient adequacy of the diet, particularly for use in settings where poverty results in monotonous diets (Martin-Prevel et al., 2017; FAO, 2021).

Like the infant and young child MDD indicator, data can be collected using a lower-burden non-quantitative recall questionnaire for food groups consumed the day before the survey. The recommended data collection methods are designed to exclude very small quantities of consumption (< 15g) (FAO, 2021).

Groups of women consuming five or more of the ten defined food groups are likely to have higher micronutrient adequacy than women consuming fewer food groups. The underlying 10-point score is also associated with micronutrient adequacy.

The 10 food groups differ slightly from the infant and young child food groups, as follows:

- Grains, white roots and tubers, and plantains
- Pulses (beans, peas, and lentils)
- Nuts and seeds
- Dairy
- Meat, poultry and fish
- Eggs
- Dark-green leafy vegetables
- Other vitamin A rich fruits and vegetables
- Other vegetables
- Other fruit

MDD-W data are increasingly available, including from the 8th phase of the DHS Program (2018–2023). Both the original MDD and the MDD-W have also been evaluated for their usefulness in predicting micronutrient adequacy for other demographic groups (for example, Ganpule-Rao et al., 2021; Diop et al., 2021).

Like the infant and young child MDD, the MDD-W was not developed to capture other characteristics of diet quality, such as NCD risk reduction. This is a key limitation, because in many resource-constrained environments, there is a growing burden of overweight, obesity and associated NCDs, following on rapid nutrition transitions (Popkin, Corvalan, & Grummer-Strawn, 2020). This limitation has been addressed in several recent efforts to develop more comprehensive indices that can still be captured using lower-burden data collection methods.

8c.10.3 Global Dietary Recommendations Score

In addition to proposing an updated HDI-2020 based on global recommendations from the WHO and the World Cancer Research Fund / American Institute for Cancer Research; (as described above), Herforth et al. (2020) also developed a Global Dietary Recommendations (GDR) Score. The GDR Score is a lower-burden index capturing adherence to these global recommendations. It was explicitly developed as a complement to the MDD-W, to fill a gap in low burden indices reflecting NCD risk reduction.

Like the MDD-W, the GDR Score is meant for use in assessing and monitoring diets at the population level, not the individual level.

Construct validity was assessed by comparing the GDR Score to the HDI-2020 in two nationally representative data sets, from Brazil and the United States, using data for ages ≥15y. Data for the GDR Score can be collected using a non-quantitative recall of food groups consumed the day before the survey. A diet quality questionnaire has been developed for this purpose (Herforth et al., 2019).

Table 8c.8 Food groups in the Global Dietary Recommendations Score, based on a one-day non-quantitative recall. Adapted from H

GDR-Healthy (positively scored)
Dark-green leafy vegetables
Vitamin A-rich orange-colored vegetables, roots, tubers
Other vegetables
Vitamin A-rich fruits
Citrus fruits
Other fruits (including red/purple/blue fruits)
Legumes
Nuts/seeds
Whole grains
GDR-Limit (negatively scored)
Sodas/sugar-sweetened beverages
Baked/grain-based sweets
Other sweets
Processed meat (double weight)
Unprocessed red meat
Deep-fried foods
Food from a fast-food restaurant, or Instant noodles
Packaged salty snacks

The GDR Score (Table 8c.8) ranges from −9 to +9, based on sets of positively and negatively scored food groups. Each food group has equal weighting except for processed meat, which has a score of −2. Overall, the score gives equal weighting to “positive” and “negative” components.

The WHO Healthy Diet guidance — one basis for the GDR Score — currently includes no recommendations on animal-source foods, though they can be good sources of bioavailable micronutrients. Processed meat and unprocessed red meat are included as negatively scored items to limit, based on World Cancer Research Fund / American Institute for Cancer Research recommendations. To capture nutrient adequacy in resource-constrained environments, the GDR Score can be considered alongside food group diversity scores such as the MDD-W, which are designed to reflect nutrient adequacy and include positive scoring for a variety of animal-source foods.

The GDR Score is a recent innovation and should be further assessed in a wider range of country contexts, with diverse cultural dietary patterns. Data to construct the GDR Score at national level will be available for 40 countries from the 2021 Gallup World Poll. These data could provide baselines for future assessments of trends. Further details and updates on the GDR Score, including country-adapted questionnaires for its measurement, are available from the website launched in 2021.

8c.10.4 Global Diet Quality Score

The Global Diet Quality Score (GDQS) represents another lower-burden approach to characterizing diet quality at population

level, including both nutrient adequacy and NCD risk (Bromage et al, 2021). In contrast to the GDR Score, the GDQS is not based on adherence to existing / current global guidance. Its development was similar to that of the AHEI, in that food groups were selected for inclusion based on a review of literature demonstrating relationships of food group intakes to outcomes. Further analyses of multiple data sets, including cross-sectional and cohort data, led to refinement of the food groups and scoring.

The GDQS evolved from an earlier index initially developed for use in the United States as a clinical screener (Rifas-Shiman et al., 2001). The global version was developed for population-level use based on extensive analyses of data sets from Africa (several countries), India, China, Mexico, and the US. Construct validity was assessed relative to nutrient adequacy, biomarkers for NCD risk, metabolic syndrome and incidence of type 2 diabetes (Bromage et al, 2021).

Data requirements for the GDQS are intermediate in complexity and, as for the MDD-W and the GDR Score, no food composition data is needed to construct the GDQS. Semi-quantitative data are needed to distinguish categories for quantity of intake for each of the 25 food group components (Table 8c.9). Two food groups (high-fat dairy and red meat), considered unhealthy when consumed in excessive amounts, have “U-shaped” scoring, such that both very low intakes and very high intakes are scored lower than intermediate intakes. Different from the other food groups in the index, which use 3 categories for quantity of intake, the high-fat dairy food group uses 4 categories to classify the quantity of intake for scoring. The “U-shaped” scoring for the high-fat dairy and red meat food groups reflects the role of these food groups both in meeting nutrient adequacy, but also in NCD risk when consumed in excess. Multiple weighting schemes were evaluated, and the selected components are not equally weighted.

Table 8c.9 Global Diet Quality Score — Food Groups and Scoring Standards. Adapted from Bromage et al (2021).

[a] *GDQS = Global Diet Quality Score; GDQS– = Global Diet Quality Score Negative Sub-metric; GDQS+ = Global Diet Quality Score Positive Sub-metric.*

[b] *Hard cheese should be converted to milk equivalents using a conversion factor when calculating total consumption of high-fat dairy for the purpose of assign*

[c] *Intakes of “High-fat dairy” > 734g/d are categorized as “Very High” and assigned zero points.*

	Categories of consumed amounts (g/d)			
	Low	**Middle**	**High**	
Included in GDQS[a] & GDQS+ metrics. Healthy food groups				
Citrus fruits	< 24	24–69	> 69	
Deep orange fruits	< 25	25–123	> 123	
Other fruits	< 27	27–107	> 107	
Dark green leafy vegetables	< 13	13–37	> 37	
Cruciferous vegetables	< 13	13–36	>36	
Deep orange vegetables	< 9	9–45	> 45	
Other vegetables	< 23	23–114	> 114	
Legumes	< 9	9–42	> 42	
Deep orange tubers	< 12	12–63	> 63	
Nuts and seeds	< 7	7–13	> 13	
Whole grains	< 8	8–13	> 13	
Liquid oils	< 2	2–7.5	> 7.5	
Fish and shellfish	< 14	14–71	> 71	
Poultry and game meat	< 16	16–44	>44	
Low-fat dairy	< 33	33–132	> 132	
Eggs	< 6	6–32	> 32	
Included in GDQS & GDQS– metrics. Unhealthy in excessive amounts				
High-fat dairy[c] (in milk equivalents)	< 35	35–142	> 142–734	

	Categories of consumed amounts (g/d)			
	Low	Middle	High	
Included in GDQS[a] & GDQS+ metrics. Healthy food groups				
Red meat	< 9	9–46	> 46	
Included in GDQS & GDQS– metrics. Unhealthy food groups				
Processed meat	< 9	9–30	> 30	
Refined grains and baked goods	< 7	7–33	> 33	
Sweets and ice cream	< 13	13–37	>37	
Sugar-sweetened beverages	< 57	57–180	>180	
Juice	< 36	36–144	> 144	
White roots and tubers	< 27	27–107	>107	
Purchased deep-fried foods	< 9	9–45	> 45	

In addition to the overall score, ranging from 0 to 49, the GDQS also has two "sub-metrics" reflecting consumption of healthy and unhealthy components (Table 8c.9). Further, two cutoffs have been identified to allow for reporting the percent of the population at high risk for poor diet quality outcomes (GDQS < 15) and the percent of the population at low risk for poor diet quality outcomes (GDQS ≥ 23).

Like the GDR Score, the GDQS is a recent innovation. Unlike the GDR Score, the semi-quantitative data required to construct the GDQS cannot be derived from non-quantitative food group recalls implemented to date in global survey programs such as the DHS, MICS and Gallup World Poll.

Lower-burden methods for the semi-quantitative recall required for the GDQS are currently under development and include a technology-assisted data collection tool and the use of visual aids (a set of 3D cubes of specified size) for collecting data to categorize the quantity of consumption per GDQS food group (Moursi et al, 2021).

Box 8c.11 summarizes key points about lower burden indices, designed for settings where detailed quantitative dietary intake data collection is infeasible.

Box 8c.11 Summary of lower burden indices

- Lower-burden indices have been developed for assessment and monitoring at population level, and not for use for individuals;
- Food group diversity indicators have been developed for several demographic groups, to proxy for micronutrient density or adequacy;
- These food group diversity indicators were not designed to reflect NCD risk;
- Recently, several new lower burden indices have been developed;
- The Global Dietary Recommendations (GDR) Score is a simple proxy indicator reflecting adherence to global dietary guidance from the World Health Organization and the World Cancer Research Fund, focused on the NCD risk;
- The Global Diet Quality Score (GDQS) reflects both nutrient adequacy and NCD risk;
- Both new scores — the GDR Score and the GDQS — were developed for adults, but may be evaluated for other groups in future;
- Components for both the GDR Score and the GDQS are food groups, and both have positive/healthy and negative/unhealthy components and sub-indices;
- Neither the GDR Score nor the GDQS requires the use of food composition data for tabulation;
- The GDR Score is based on a non-quantitative food group recall data and the GDQS is based on semi-quantitative data.

This page titled 7.10: Lower-burden diet quality indices (8c.10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)

Current food systems are now known to contribute a significant share of global greenhouse gas emissions and to contribute to land conversion, deforestation, and biodiversity loss; agriculture also accounts for the majority of global freshwater withdrawals (FAO & WHO, 2019). Recognition of this has led to efforts to define diets that are both healthy and environmentally sustainable, and to calls for integration of sustainability considerations in national food-based dietary guidelines, as noted above (Gonzalez Fischer & Garnett, 2016; Springmann et al., 2020).

Table 8c.10 EAT-Lancet healthy reference diet, for an intake of 2500kcals per day. Adapted from W. Willett et al. (2019)

[a] *Wheat, rice, dry beans and lentils are dry, raw.*

[b] *Consists of fish and shellfish, including both wild and farmed.*

[c] *Includes olive, soy bean, rapeseed, sunflower, and peanut oil.*

Component	Macronutrient intake (possible range), g/d
Whole grains[a]	
Rice, wheat, corn and other	232 (total grains 0–60% of energy)
Tubers or starchy vegetables	
Potatoes and cassava	50 (0–100)
Vegetables	
All vegetables	300 (200–600)
Dark green vegetables	100
Red and orange vegetables	100
Other vegetables	100
Fruits	
All fruit	200 (100–300)
Dairy foods	
Whole milk or derivative equivalents (e.g. cheese)	250 (0–500)
Protein sources	
Beef and lamb	7 (0–14)
Pork	7(0–14)
Chicken and other poultry	29 (0–58)
Eggs	13 (0–25)
Fish[b]	28 (0–100)
Dry beans, lentils, and peas[a]	50 (0–100)
Soy foods	25 (0–50)
Peanuts	25 (0–75)
Tree nuts	25
Added fats	

Component	Macronutrient intake (possible range), g/d
Palm oil	6.8 (0–6.8)
Unsaturated oils[c]	40 (20–80)
Lard or tallow	5 (0–5)
Added sugars	
All sweeteners	31 (0–31)

In 2019, the EAT-Lancet Commission(Willet et al., 2019) addressed sustainability concerns and provided a global "healthy reference diet", with quantitative targets (and ranges) for food group intake, based on a 2500 calorie diet (Table 8c.10).The reference diet was intended to be flexible and adaptable to various cultural contexts. The EAT-Lancet reference diet is high in plant-source foods and with limited amounts of animal-source foods, with the low end of the "healthy" range for these foods set at zero. Currently, there is no established global guidance on the appropriate balance of plant-source and animal-source foods.

Some have raised concerns that the EAT-Lancet reference diet unnecessarily restricts nutrient-dense animal-source foods (Raiten et al., 2020). Vaidyanathan(2021) summarizes some of the controversies and concerns. For nutrition-insecure women in low- and middle-income countries, Hanley-Cook et al.(2021) demonstrated that micronutrient adequacy improved when the EAT-Lancet intake ranges were modified by imposing non-zero minimum quantities for nutrient-dense animal-source food groups.

Meantime, several groups have proposed indices for healthy, sustainable diets either based on national guidance (Harray et al., 2015) or based on the EAT-Lancet reference diet (Knuppel et al., 2019; Trijsburg et al., 2019). Stubbendorff et al., 2021).Given this is a new area of inquiry, it is likely that additional research on sustainable healthy diets will yield additional indices and measurement tools in coming years.

This page titled 7.11: Incorporating sustainability in diet quality definitions and indices (8c.11) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

7.12: Summary of diet quality indices (8c.12)

This section has presented a wide variety of diet quality indices developed for diverse uses. Indices are closely related to definitions of diet quality and have evolved with changing definitions. At national level, food-based dietary guidelines describe healthy diets, and indices developed to capture adherence to such guidelines are widely used. Currently, many indices — whether based on national, global, or other guidance — aim to reflect both nutrient adequacy and NCD risk reduction. Concerns with the impact of diets on planetary health are growing, and this is likely to result in new and more comprehensive indices in the near future.

Indices also vary widely in the data required for their calculation. Most indices developed for use in high-income data-rich environments require both detailed quantitative dietary intake data, and associated resources such as food composition databases. However, indices with lower data requirements have recently been proposed, for global use.

This page titled 7.12: Summary of diet quality indices (8c.12) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

8: INTRODUCTION TO ANTHROPOMETRY (CHAPTER 9)

Abstract

Anthropometric measurements are used to assess body size and body composition. The measurements are simple, safe, and non-invasive and provide information on past exposure, but cannot detect short-term disturbances or deficiency of a specific nutrient. There are three major sources of error in anthropometry: (i) measurement errors, (ii) alterations in the composition and physical properties of certain tissues, and, (iii) use of invalid assumptions in the derivation of body composition from the measurements.

Anthropometric indices are derived from two or more raw measurements and are essential to interpret and group the anthropometric data. Selection of indices must take into account their sensitivity, specificity, predictive value, and any potential modifying factors.Examples of indices include weight-for-height, body mass index (weight kg) / (height m)^2, and waist-hip circumference ratio. Anthropometric indices are often evaluated by comparison with predetermined reference limits or cutoff points. Calculation of the number and proportion of individuals (as %) with anthropometric indices below or above a designated reference limit or cutoff, generates "anthropometric indicators" that can be used in clinical and public health settings to classify individuals at risk of malnutrition. Examples of indicators used in this way include mid-upper-arm circumference (MUAC) with a cutoff < 115mm to identify severe acute malnutrition (SAM) in children 6–60mos, and WHZ < −2, BMIZ > +2, and HAZ < −2, used by WHO and UNICEF to define wasting, overweight, and stunting respectively in children < 5y and to define prevalence thresholds and identify priority countries. Cutoffs, unlike statistically derived reference limits, are based on functional impairment or clinical signs of malnutrition, and occasionally mortality.

The reference growth data recommended by WHO for international use are the prescriptive WHO Child Growth Standards for 0–5y, and the WHO growth reference data for older 5–19y. Updated childhood growth charts are also available for U.S. infants age 0–36mos and children 2–20y. Local reference data are preferred for body composition, although few are available. Instead, WHO recommends using reference data for MUAC, triceps, and subscapular skinfolds collected for the WHO Child Growth Standards.

The term "nutritional anthropometry" first appeared in "Body Measurements and Human Nutrition" (Brožek, 1956) and was later defined by Jelliffe (1966) as:

> "measurements of the variations of the physical dimensions and the gross composition of the human body at different age levels and degrees of nutrition"

Subsequently, a number of publications made recommendations on specific body measurements for characterizing nutritional status, standardized measurement techniques, and suitable reference data (Jelliffe, 1966; WHO, 1968; Weiner and Lourie, 1969). Today, anthropometric measurements are widely used for the assessment of nutritional status and health, at both the individual and population levels. One of their main advantages is that anthropometric measurements may be related to past exposures, to present processes, or to future events (WHO, 1995).

For individuals in low-income countries, anthropometry is particularly useful when there is a chronic imbalance between intakes of energy, protein, and certain micronutrients. Such disturbances modify the patterns of physical growth and the relative proportions of body tissues such as fat, muscle, and total body water. For individuals in clinical settings, anthropometry can be used to diagnose failure to thrive in infants and young children, and monitor overweight and obesity in children and adults.

At the population level, anthropometry has an important role in targeting interventions through screening, in assessing the response to interventions, in identifying the determinants and consequences of malnutrition, and in conducting nutritional surveillance. Increasingly, anthropometry is also being used to characterize and compare the health and nutritional status of populations across countries (WHO/UNICEF, 2019).

8.1: Measurements, indices, and indicators

8.2: Advantages and limitations of anthropometry

8.3: Errors in anthropometry
8.4: Interpretation and evaluation of anthropometric data

This page titled 8: Introduction to Anthropometry (Chapter 9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Measurements, indices, and indicators

Anthropometric measurements are of two types. One group of measurements assesses body size, the other group appraises body composition. The most widely used measurements of body size are stature (length or height), weight, and head circumference; see Chapter 10 for more details. The anthropometric measurements of body composition are based on the classical “two component model” in which the body is divided into two major compartments, fat mass and the fat free mass. Skinfold thickness measurements are used to estimate of the size of the subcutaneous fat depot, which, in turn provides an estimate of total body fat: over one third of total body fat is estimated to be subcutaneous fat. The distribution of body fat is also important, with the measurement of waist circumference used increasingly as a proxy for the amount of intra-abdominal visceral fat. Waist circumference is recommended for use in population studies (WHO, 2011), as well as in clinical practice for the evaluation and management of patients with overweight or obesity (Ross et al., 2020).

The fat-free mass consists of the skeletal muscle, non-skeletal muscle, soft lean tissue, and the skeleton. A major component of the fat-free mass is body muscle. As this is composed of protein, assessment of muscle mass can provide an indirect assessment of the protein reserves of the body. Measurements of thigh circumference and mid-upper-arm circumference (MUAC) can be used to assess skeletal muscle mass (Müller et al., 2016).Measurement of MUAC is especially useful for young children < 5y in emergency settings such as famines and refugee crises. In such settings, children often have a small amount of subcutaneous fat, so changes in MUAC tend to parallel changes in muscle mass; see Chapter 11 for more details.

Anthropometric indices are usually calculated from two or more raw measurements, and are essential for the interpretation and grouping of measurements collected in nutritional assessment. For example, the measurement of a child's body weight is meaningless unless it is related to the age or height of a child. In young children the three most commonly used growth indices are weight-for-age, height-for-age, and weight-for-height. The first two indices reflect body weight or height relative to chronological age, whereas weight-for-height assesses body weight relative to height.

Body mass index (BMI) is also widely used in children and adults to assess underweight, overweight, and obesity, and is calculated as (weight kg) / (height m)^2. When height cannot be measured, as may occur in bed-bound or frail individuals, published equations based on a range of body measurements such as knee height, lower leg length, arm span, and ulna length can be used to provide an approximate estimate of height. Examples of equations for estimating height from these body measurements in adults are given in Madden et al. (2016). However, their usefulness for hospitalised patients may be questionable, if the equations have been derived from young and healthy populations (Reidlinger et al., 2014).

Examples of body composition indices include a combination of triceps skinfold and mid-upper-arm circumference, which together can be used to estimate mid-upper-arm fat area and mid-upper-arm muscle circumference or area, surrogates for total body fat content, and muscle mass, respectively. Other measurement combinations include the waist-hip ratio (i.e., the waist circumference divided by the hip circumference), an additional index of the distribution of body fat which can be measured more precisely than skinfolds. Moreover, measurements of waist-hip ratio as a surrogate for abdominal obesity, appear to be a stronger independent risk factor for risk of myocardial infarction, stroke and premature death than BMI, especially among men (Larsson et al., 1984; Lapidus et al., 1984).

In an effort to obtain more reliable estimates of percentage body fat and fat-fat-free mass based on anthropometric measurements in healthy adults, the sum of skinfold thickness measurements from multiple anatomical sites is also used in conjunction with population-specific or generalized regression equations to predict body density, and in turn, the percentage of body fat using one of three empirical equations. Once the percentage of body fat is calculated, total body fat content and the fat-free mass can be derived (see Chapter 11 for more details). Again, many of the prediction equations were developed on young, healthy, lean Caucasian population groups and, hence, are less appropriate for malnourished, obese, or elderly subjects or for other racial groups.

Anthropometric indices are often evaluated by comparison with the distribution of appropriate anthropometric reference data using standard deviation scores (Z-scores) or percentiles. (see Section 9.4.3). From this, the number and proportion of individuals (as %) with anthropometric indices below or above a predetermined reference limit or cutoff are often calculated. A commonly used reference limit for the three main growth indices is a Z-score of –2 (i.e., below the WHO reference median) (Section 9.4.2). When used in this way, the index and its associated reference limit or cutoff become an “indicator”; these are discussed below.

Anthropometric indicators are constructed from anthropometric indices, with the term “indicator” relating to their use in nutritional assessment, often for public health, or socio-medical decision-making at the population level. Indicators are also used in clinical settings to identify individuals at risk of malnutrition. To be valid, a substantial proportion of the variability of an anthropometric indicator should be associated with differences in nutritional status. WHO (1995) provide a detailed classification of recommended anthropometric indicators based on their uses for both targeting and assessing response to interventions, identifying determinants of malnutrition, or predicting malnutrition in populations of infants and children.

Anthropometric indicators should be chosen carefully in relation to both their proposed use and their attributes. Indicators vary in their validity, sensitivity, specificity, and predictive value; these characteristics are discussed briefly in Section 9.4.3. For example, although the indicator weight-for-age < –2 Z-score is still widely used in health centers in many low-income countries for screening young children at risk of malnutrition, it is inappropriate. Children who are stunted but of normal weight, or alternatively, tall and thin may be incorrectly diagnosed as “healthy”. Instead, in these countries, the indicator length/height-for-age < –2 Z-score should be used (Ruel et al., 1995).

Further, several factors will affect the magnitude of the expected response of an anthropometric indicator. These may include the degree of deficiency, age, sex, and physiological state of the target group. Some examples of frequently used anthropometric indicators and their corresponding application are shown in Table 9.1.

Table 9.1 Anthropometric indicators and their corresponding applications.

Anthropometric indicator	Application
Proportion of children (of defined age and sex) with WHZ < –2	Prevalence of wasting
Proportion of children (of defined age and sex) with HAZ < –2	Prevalence of stunting
Proportion of children (of defined age and sex) with WAZ < –2	Prevalence of underweight
Proportion of children 0–5y (of defined age and sex) with BMIZ > +2 or BMIZ > +3	Prevalence of overweight or obesity
Proportion of adult women or men with waist-hip ratios > 0.85 (F) and > 0.90 (M)	Prevalence of abdominal obesity and thus risk of metabolic syndrome
Proportion of children 6–60mos with MUAC < 115mm	Prevalence of severe acute malnutrition (SAM)
Proportion of children with SAM who have MUAC > 125mm and no edema for at least 2wk after receiving treatment for SAM	Prevalence of children ready for discharge following treatment for SAM

This page titled 8.1: Measurements, indices, and indicators is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

8.2: Advantages and limitations of anthropometry

Anthropometric measurements are of increasing importance in nutritional assessment as they have many advantages (Box 9.1). However, anthropometric measures are relatively insensitive and cannot detect disturbances in nutritional status over short periods of time. Furthermore, nutritional anthropometry cannot identify any specific nutrient deficiency and, therefore, is unable to distinguish disturbances in growth and body composition induced by nutrient deficiencies (e.g., zinc) from those caused by imbalances in protein and energy intake.

Certain non-nutritional factors (such as disease, genetic influences, diurnal variation, and reduced energy expenditure) can lower the specificity and sensitivity of anthropometric measurements (Section 1.4), although such effects generally can be excluded or taken into account by appropriate sampling and experimental design.

Nevertheless, nutritional anthropometry can be used to monitor changes in both growth and body composition in individuals (e.g., hospital patients) and in population groups, provided sources of measurement error and the effects of confounding factors are minimized (Ulijaszek & Kerr, 1999).

Box 9.1. The advantages of anthropometry measurements in nutritional assessment

- **Simple, safe, noninvasive techniques** are involved, which can be used at the bedside of a single patient, but are also applicable to large sample sizes.
- **Inexpensive equipment** is required. It is portable, and durable and can be made or purchased locally.
- **Relatively unskilled personnel** can perform the measurement procedures if adequately trained
- **Methods can be precise and accurate,** if standardized techniques and trained personnel are used.
- **Retrospective information** is generated on past long-term nutritional history, which cannot be obtained with equal confidence using other techniques.
- **Mild to moderate undernutrition,** as well as severe states of under- or overnutrition, can be identified.
- **Changes in nutritional status** over time and from one generation to the next, a phenomenon known as the secular trend, can be evaluated.
- **Screening tests** that identify individuals at high risk to under- or overnutrition can be devised.

This page titled 8.2: Advantages and limitations of anthropometry is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

8.3: Errors in anthropometry

Errors can occur in nutritional anthropometry which may affect the precision, accuracy, and validity of the measurements, and thus indices and indicators. Three major sources of error are significant: measurement errors, alterations in the composition and physical properties of certain tissues, and the use of invalid assumptions in the derivation of body composition from anthropometric measurements (Heymsfield and Casper, 1987).

Measurement errors arise from examiner error resulting from inadequate training, instrument error, and difficulties in making the measurement (e.g., skinfold thicknesses). The major sources of measurement error in anthropometry are shown in Boxes 9.2 and 9.3. Both random and systematic measurement errors may occur which reduce the validity of the index and any indicator constructed from the index; they have been extensively reviewed by Ulijaszek and Kerr (1999).

Box 9.2: Common errors and possible solutions when measuring length, height, and weight.

All measurements

- **Inadequate instrument:** Select method appropriate to resources.
- **Restless child:** Postpone measurement or involve parent in procedure or use culturally appropriate procedures.
- **Errors in reading equipment:** Training and refresher exercises, stressing accuracy, with intermittent oversight by supervisor.
- **Errors in recording results:** Record results immediately after measurement and have record checked by a second person.

Length

- **Incorrect method for age:** Use length only when child is < 2y.
- **Footwear/headwear not removed:** Remove as local culture permits (or make allowances).
- **Head not in correct Frankfurt plane:** Correct position of child before measuring.
- **Child not straight along board and/or feet not parallel with movable board:** Have assistant and child's parent present; don't take the measurement while the child is struggling; settle child.
- **Board not firmly against heels:** Correct pressure should be practiced.

Height

- **Incorrect method for age:** Use only when child is ≥ 2y.
- **Footwear/headware not removed;** Remove as local culture permits (or make allowances).
- **Head not in correct Frankfurt plane, subject not straight, knees bent, or feet not flat on floor:** Correct technique with practice and retraining; provide adequate assistance; calm noncooperative children.
- **Board not firmly against head:** Lower head board to compress hair.

Weight

- **Room cold, no privacy:** Use appropriate clinic facilities.
- **Scale not calibrated to zero:** Re-calibrate after every subject.
- **Subject wearing heavy clothing:** Remove or make allowances for clothing.
- **Subject moving or anxious as a result of prior incident:** Wait until subject is calm or remove the cause of anxiety.

9.3.1 Random measurement errors and precision

Random measurement errors limit precision or the extent to which repeated measurements of the same variable give the same value. Random measurement errors can be minimized by training personnel to use standardized techniques and precise, correctly calibrated instruments (Lohman et al., 1988) . Furthermore, the precision (and accuracy) of each measurement technique should be firmly established prior to use. To improve precision, two or three measurements on each individual should be conducted.

A description of the measurement techniques used in the WHO Multicenter Growth Reference Study (MGRS) are available in de Onis et al. (2004), as well as in an anthropometric training video from WHO. In the WHO MGRS the equipment was calibrated regularly, using standard weights over the full weight range for the portable electronic weighing scales, metal rods of known length for both the infantometer and stadiometer, and calibration blocks of varying widths for the skinfold calipers.

Poor precision often reflects within-examiner error, but between-examiner error may also be significant in surveys with multiple examiners. The precision of a measurement technique can be assessed by calculating:

- Technical error of the measurement (TEM)
- percentage technical error (%TEM)
- Coefficient of reliability (R)

These parameters can be calculated for each anthropometric measurement technique from repeated measurements on each subject made within a few minutes to avoid physiological fluctuations. A minimum of 10 subjects is recommended.

Box 9.3: Common errors and possile solutions when measuring mid-upper-arm circumference, head circumference, and triceps skinfold.

It is particularly important with these measurements to use the correct techniques. This requires training, supervision, and regular refresher courses. Always take into account any cultural problems, such as the wearing of arm bands etc.

Arm circumference

- **Subject not standing in correct position:** Position subject correctly.
- **Tape too thick, stretched, or creased:** Use correct instrument.
- **Wrong arm:** Use left arm.
- **Mid-arm point incorrectly marked:** measure and remark midpoint carefully.
- **Arm not hanging loosely by side during measurement:** ask subject to allow arm to hang loosely.
- **Examiner not comfortable or level with subject:** position subject correctly relative to examiner.
- **Tape around arm not at midpoint:** reposition tape.
- **Tape too tight (causing skin contour indentation):** loosen tape.
- **Tape too loose:** carefully tighten tape.

Head circumference

- **Occipital protuberance / supraorbital landmarks poorly defined:** position tape correctly.
- **Hair crushed inadequately:** carefully tighten tape.
- **Ears under tape, or tension position poorly maintained at time of reading:** repeat after positioning tape correctly.
- **Headwear not removed:** remove as local culture permits.

Triceps fatfold

- **Wrong arm:** use left arm.
- **Mid-arm point or posterior plane incorrectly measured or marked:** measure and remark midpoint carefully.
- **Arm not loose by side during measurement:** ask subject to allow arm to hang loosely.
- **Finger-thumb pinch or caliper placement too deep (muscle) or too superficial (skin):** correct technique with training and supervision.
- **Caliper jaws not at marked site; reading done too early, pinch not maintained, caliper handle not released:** correct technique with training and supervision.
- **Examiner not comfortable or level with subject:** ensure examiner is correctly positioned.

From: (Zerfas AJ, 1979)

The technical error of the measurement (TEM) is the square root of the measurement error variance. TEM is expressed in the same units as that of the anthropometric measurement under study and is often age dependent. The value is also related to the anthropometric characteristics of the study group. The calculation varies according to the number of replicate measurements made. For one examiner making two measurements, TEM =

$$\sqrt{(\Sigma D^2)/2\ N}$$

where D = the difference between two measurements and N = number of subjects. For more than two measurements, the equation is more complex, and TEM =

$$\sqrt{\sum_{1}^{N}\left[\left(\sum_{1}^{K}M^2\right)-\left(\left(\sum_{1}^{K}M\right)^2/K\right)\right]/[N(K-1)]}$$

where N = number of subjects, K is the number of determinations of the variable taken on each subject, and M_n is the nth replicate of the measurement, where *n* varies from 1 to K.

Table 9.2: shows the calculation of TEM from measurements of stature performed four times on 10 subjects by a single anthropometrist.

Table 9.2 Sample calculation of technical error of measurement (TEM) from repeat measurements of stature (m) carried out by one anthropometrist on 10 subjects. M= measurement, K= number of replicates, N= number of subjects.

Subject	Stature (m) as determined on repeat				(1)	(2)	Diff.
no.	1	2	3	4	ΣM^2	$(\Sigma M)^2/K$	(1) – (2)
1	0.865	0.863	0.863	0.864	2.984259	2.984256	0.000003
2	1.023	1.023	1.027	1.025	4.198412	4.198401	0.000011
3	0.982	0.980	0.989	0.985	3.873070	3.873024	0.000046
4	0.817	0.816	0.812	0.817	2.660178	2.660161	0.000017
5	0.901	0.894	0.900	0.903	3.236446	3.236401	0.000045
6	0.880	0.876	0.881	0.881	3.094098	3.094081	0.000017
7	0.948	0.947	0.947	0.946	3.587238	3.587236	0.000002
8	0.906	0.905	0.907	0.908	3.286974	3.286969	0.000005
9	0.924	0.924	0.926	0.924	3.418804	3.418801	0.000003
10	0.969	0.987	1.002	0.993	3.942343	3.942210	0.000133
							Σ = 0.000282
TEM = = = 0.00307							

Note that the size of the measurement also influences the size of the associated TEM, so that comparisons of precision of different anthropometric measurements using TEM cannot be made easily. This is highlighted in Table 9.3 in which the TEM for five anthropometric measurements taken during the initial standardization session conducted at the Brazilian site of the WHO Multicentre Growth Reference Study (MGRS) are presented (de Onis et al., 2004). Table 9.3 also depicts the maximum allowable differences between the measurements of two observers that were used in the WHO MGRS, and set based on TEMs achieved during the standardization session.

Table 9.3. Maximum allowable differences between the measurements of two observers.
TEM: Technical error of measurement
From: de Onis et al., 2004.

measurement	Brazil TEM from pilot study	Maximum allowa
Weight	Not available	100g
Length	2.5mm	7.0mm
Head circumference	1.4mm	5.0mm
Arm circumference	1.8mm	5.0mm

measurement	Brazil TEM from pilot study	Maximum allowa
Triceps skinfold	0.44mm	2.0mm
Subscapular skinfold	0.43mm	2.0mm

Percentage TEM has been recommended to overcome the difficulty of the TEM being dependent on the size of the original measurement (Norton and Olds, 1996). The percentage technical error of the measurement is analogous to the coefficient of variation and is calculated as:

$$\%\mathrm{TEM} = (\mathrm{TEM}/\ \mathrm{mean}\ ) \times 100\%$$

Note that %TEM has no units and can be used to make direct comparisons of all types of anthropometric measurements. It cannot be used, however, when more than one examiner is involved, as then both within- and between-examiner errors are involved. Ulijaszek and Kerr (1999) describe ways to deal with this more complex case.

The coefficient of reliability (R) is an alternative approach that is widely used for comparing measurement errors among anthropometric measurements. It ranges from 0 to 1 and can be calculated using the following equation:

$$\mathrm{R} = 1 - \left((\mathrm{TEM})^2/\mathrm{s}^2\right)$$

where s^2 is the between-subject variance. The coefficient indicates the proportion of between-subject variance in a measured population which is free from measurement error. Hence, a measurement with R = 0.95 indicates that 95% of the variance is due to factors other than measurement error.

Whenever possible, a coefficient of reliability > 0.95 should be sought. Coefficients of reliability can be used to compare the relative reliability of different anthropometric measurements, and the same measurements in different age groups, as well as for calculating sample sizes in anthropometric surveys.

More details of standardization procedures and calculation of precision using TEM, percentage TEM, and coefficient of reliability are given in Lohman et al. (1988). In general, the precision of weight and height measurements are high. However, for waist and hip circumferences, between-examiner error tends to be large and it is preferable for only one examiner to take these measurements. Because skinfolds are notoriously imprecise, both within- and between-examiner errors can be large. Therefore, rigorous training using standardized techniques and calibrated equipment are critical when skinfold measurements are taken.

9.3.2. Systematic measurement errors and accuracy

Systematic measurement errors affect the accuracy of anthropometric measurements or how close the measurements are to the true value. The most common form of systematic error in anthropometry results from equipment bias. For example, apparent discrepancies in skinfold measurements performed on the same person but with different calipers may be due to compression differences arising from variations in spring pressure and surface area of the calipers (Schmidt & Carter, 1990); Harpenden and Holtain skinfold calipers consistently yield smaller values than Lange calipers(Gruber et al., 1990). Errors arising from bias reduce the accuracy of the measurement by altering the mean or median value, but have no effect on the variance. Hence, such errors do not alter the precision of the measurement.

The timing of some anthropometric measurements of body size and composition is also known to be critical, particularly for short-term growth studies: progressive decreases in the height of an individual during the day as a consequence of compression of the spinal column, for example, may seriously compromise the accuracy of height velocity measurements.

The determination of accuracy in anthropometry is difficult because the correct value of any anthropometric measurement is never known with absolute certainty. In the absence of absolute reference standards, the accuracy of anthropometric measurements is estimated by comparing them with those made by a criterion anthropometrist (Ulijaszek & Kerr, 1999), a person who has been highly trained in the standardized measurement techniques and whose measurements compare well to those from another criterion anthropometrist.

In preparation for the compilation of the new WHO Child Growth Standard, four anthropometrists were trained and standardized against a criterion anthropometrist, designated as the “lead” anthropometrist; see de Onis et al. (2004) for more details. All anthropometric measurements were taken and recorded independently by two designated anthropometrists, and their measurement values compared for maximum allowable differences (Table 9.3). Targets for sports anthropometrists are also available (Gore

et al., 1996).

Attempts should always be made to minimize measurement errors. In longitudinal studies involving sequential anthropometric measurements on the same group of individuals (e.g., surveillance), it is preferable, whenever possible, to have one person carrying out the same measurements throughout the study to eliminate between-examiner errors. This is particularly critical when increments in growth and body composition are calculated; such increments are generally small and are associated with two error terms, one on each measurement occasion. Recommendations for the minimum intervals necessary to provide reliable data on growth increments during infancy and early childhood (Guo et al., 1991) and adolescence(WHO, 1995) are available.

In large regional surveillance studies, it is often necessary to use several well-trained anthropometrists. In such circumstances, the between-examiner differences among anthropometrists must be monitored throughout the study to maintain the quality of the measurements and thereby to identify and correct systematic errors in the measurements. This practice was followed during the WHO MGRS (de Onis et al., 2004).

In studies involving two longitudinal measurements, the TEM can be calculated to estimate the proportion of the difference that can be attributed to measurement error. For example, with a TEM of 0.3 for a given anthropometric measurement, the TEM for the difference between two measurements is:

$$\sqrt{(0.3)^2 + (0.3)^2} = 0.42$$

because both TEM values contribute to the variance in the difference. Only if the difference exceeds 2 × 0.42 = 0.84 is there a 95% probability that the difference exceeds the measurement error alone.

Once assured that such differences are not a function of measurement error, then any changes in growth and body composition can be correlated with factors such as age, the onset of disease, response to nutrition intervention therapy, and so on.

The collection of longitudinal anthropometric data is more time consuming, expensive, and laborious than from cross-sectional surveys, and, as a result, the sample size is generally smaller. Hence, the probability of systematic sampling bias (Section 1.4.2) is generally greater than in more extensive cross-sectional surveys.

For cross-sectional studies, the examiners should be rotated among the subjects to reduce the effect of measurement bias of the individual examiners. Statistical methods exist for removing anthropometric measurement error from cross-sectional anthropometric data; details are given in Ulijaszek and Lourie (1994).

Cross-sectional surveys are useful for comparing population groups, provided that probability sampling techniques have been used to ensure that the samples are representative of the populations from which they are drawn (Chapter 1). Recently WHO has provided countries with tools to develop or strengthen their surveillance systems so they have the capacity to monitor changes in the Global Nutrition Targets for 2030. They include the following anthropometric indicators: stunting, wasting, low birthweight, and childhood overweight. For more details see: WHO Nutrition Tracking Tool.

9.3.3 Errors from changes in tissue composition and properties

Variation in the composition and physical properties of certain tissues may occur in both healthy and diseased subjects, resulting in inaccuracies in certain anthropometric measurements. Even among healthy individuals, body weight may be affected by variations in tissue hydration with the menstrual cycle (Heymsfield and Casper, 1987; Madden and Smith, 2016).

Skinfold thickness measurements may be influenced by variations in compressibility and skin thickness with age, gender, and the level of tissue hydration (Martin et al., 1992; Ward and Anderson, 1993). For example, repeated measurements of skinfolds, over a short period (i.e., 5min), may actually decrease accuracy of skinfolds because later measurements are more compressed due to the expulsion of water from the adipose tissue at the site of the earlier measurement (Ulijaszek and Kerr, 1999).

The accuracy of waist circumference is affected by both the phase of respiration at the point of measurement and by the tension of the abdominal wall. The phase of respiration is important because it determines the extent of fullness of the lungs and the position of the diaphragm at the time of the measurement. Increasing the tension of the abdominal wall (by sucking in) is frequently an unconscious reaction which is also important because it reduces the waist measurement. To minimize these errors, WHO (2011) recommends advising the subject to relax and take a few deep, natural breaths before the actual measurement and at the end of normal expiration.

In addition, during aging, demineralization of the bone and changes in body water may result in a decrease in the density of the fat-free mass (Visser et al., 1994; JafariNasabian et al., 2017), which are not always taken into account when calculating total body fat and hence fat-free mass from skinfolds via body density (see Chapter 11 for more details).

9.3.4 Invalid models and errors in body composition

Invalid assumptions may lead to erroneous estimates of body composition when these are derived from anthropometric measurements, especially in obese or elderly patients and those with protein-energy malnutrition or certain disease states. For instance, use of skinfold thickness measurements to estimate total body fat assumes that (a) the thickness of the subcutaneous adipose tissue reflects a constant proportion of the total body fat and (b) the sites selected represent the average thickness of the subcutaneous adipose tissue. In fact, the relationship between subcutaneous and internal fat is nonlinear and varies with body weight, age, and disease state. Very lean subjects have a smaller proportion of body fat deposited subcutaneously than do obese subjects, and in malnourished persons there is probably a shift of fat storage from subcutaneous to deep visceral sites. Variations in the distribution of subcutaneous fat also occurs with age, sex, and ethnicity or race (Wagner and Heyward, 2000; He et al., 2002).

Estimates of mid-upper-arm muscle area are used as an index of total body muscle and the fat-free mass (Chapter 11), regardless of age and health status of the subjects. Such estimates are made, despite the known changes in the relationship between arm muscle and fat-free mass with age and certain disease states (Heymsfield and McManus, 1985), and the questionable accuracy of the algorithms used (Martine et al., 1997). Moreover, even the corrected algorithms developed for adults overestimate arm muscle area in obese persons when compared with the determination by computerized tomography (Forbes et al., 1998).

Increasingly, body composition is assessed by laboratory methods; these are described in Chapter 14. Even laboratory methods are based on certain assumptions that have been challenged in recent years. For example, until recently, densitometry, frequently using underwater weighing, has been the gold standard reference method for the determination of the percentage of body fat. The assumptions used in densitometry are that the densities of the fat mass and fat-free mass are constant at 0.90 and 1.10kg/L, respectively (Chapter 11). Several researchers have questioned the validity of using a constant density of the fat-free mass for groups who vary in age, gender, levels of body fatness, and race or ethnicity (Visser et al., 1997). During aging, the density of the fat-free mass may decrease due to demineralization of the bone and changes in body water, as noted above (Visser et al., 1994; JafariNasabian et al., 2017), which are not always taken into when calculating total body fat from skinfolds via body density, leading to a1%–2% overestimate of the body fat content in such subjects (Deurenberg et al., 1989); see Chapter 11 for more details.

In contrast, persons of African descent have a larger fat-free mass because they have a greater bone mineral density and body protein content compared to Caucasians (Wagner and Heyward, 2000). Such differences lead to an underestimate of body fat, when generalized equations developed for Caucasians are used.

Percentage of body fat can also be determined using an isotope dilution technique and dual-energy X-ray absorptiometry (DXA) (Chapter 14). Both of these methods assume a constant hydration of the fat-free mass (i.e., 73.2% water content), despite knowledge that it varies with age (Wang et al., 1999), obesity, and pregnancy (Hopkinson et al., 1997), and throughout the course of a clinical condition (e.g., inflammation) (Müller et al., 2016). When the actual hydration of fat-free mass is higher than the assumed value, then the percentage of body fat is underestimated by isotope dilution techniques (Chapter 14) (Deurenberg-Yap et al., 2001). For example, even when pregnancy-specific values for hydration have been applied to account for the increased accretion of water that occurs during pregnancy, individual estimates of fat mass using isotope dilution differed by > 3kg from values based on the four-compartment model (Hopkinson et al., 1997). In contrast, hydration effects on estimates of fat by DXA are not significant (Pietrobelli et al., 1998).

Fortunately, the advent of multicomponent models (i.e., the 4-compartment-model) with minimal assumptions for assessing body composition circumvent the use of older methods, which use assumptions that are not always valid for certain ethnic groups or the elderly (Müller et al., 2016). Nevertheless, the use of multicomponent models is expensive, requiring more time and facilities.

This page titled 8.3: Errors in anthropometry is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

8.4: Interpretation and evaluation of anthropometric data

Anthropometric indices are derived from two or more raw measurements, as noted earlier. Normally, it is these indices that are interpreted and evaluated — not the raw measurements. Anthropometric indices can be used at both the individual and the population levels to assess nutritional status and to screen and assess a response during interventions. In addition, in populations, anthropometry can be used to identify the determinants and consequences of malnutrition and for nutritional surveillance. To achieve these objectives, knowledge of factors that may modify or "condition" the interpretation of abnormal anthropometric indices is generally required. These conditioning factors are briefly discussed below, together with anthropometric reference data and methods to evaluate anthropometric indices, including classification systems that identify individuals and populations as "at risk" for malnutrition.

9.4.1 Conditioning factors

A variety of factors are known to modify or condition the interpretation of anthropometric data and must be taken into account. Some important examples include age, birthweight, birth length, gestational age, sex, parental stature, and feeding mode during infancy. Maturation during adolescence, prepregnancy weight, maternal height, parity, smoking, pregnancy, and ethnicity are major conditioning factors for adults (WHO, 1995).

Information on some of these conditioning factors can be obtained by physical examinations, questionnaires, or self-reports. An accurate assessment of age is especially critical for the derivation of many anthropometric indices used to identify abnormal anthropometry, notably height-for-age and weight-for-age. See: WHO Child Growth Standards.

Age is also important for categorizing the data into the age groups recommended by WHO for analysis and interpretation (WHO/UNICEF, 2019); see Chapter 13 for details. In more affluent countries, the assessment of age using birth certificates is generally easy, but in some low-income countries, local calendars of special events are often constructed to assist in identifying the birth date of a child.

Alternatively, for young children, age is sometimes assessed by counting deciduous teeth. This method is most appropriate at the population level because of the wide variation among individuals in the timing of deciduous eruption (Delgado et al., 1975). For individuals, bone age can be estimated from the left-hand-wrist radiograph using the Tanner Whitehouse II method (Tanner et al., 1983). Gorstein (1989) has highlighted the marked discrepancies that may occur in the prevalence estimates of undernutrition during infancy when different methods are used to determine age.

With infants, an accurate assessment of birth weight, and, if possible, birth length and gestational age, is also important (Hediger et al., 1999). Assessment of gestational age is especially critical for the interpretation of both size-for-age measurements during infancy and the neurodevelopmental progress of preterm infants. It is also essential for the management of pregnancy and the treatment of new-born infants.

Several strategies are available for estimating gestational age. Prenatal measures of gestational age include calculating the number of completed weeks since the beginning of the last menstrual period, prenatal ultrasonography, and clinical methods; they are all described in Chapter 10. In public health settings, the definition of gestational age on the basis of the last menstrual period is most frequently used, although it is associated with several problems: errors may occur because of irregular menses, bleeding early in pregnancy, and incorrect recall by mothers. Prenatal ultrasonography during the first or second trimester, although considered the gold standard method for assessment of gestational age, is not universally available, especially in low-income countries. Furthermore, the quality of both the equipment used and the technical training varies.

For studies of the adolescent age group, defined by WHO (1995) as 10–19y, information on maturation should also be collected in view of the marked variation in the timing of the maturational changes during adolescence. The best measure of maturity is bone age — often termed skeletal maturation — because it can be obtained for both sexes over a wide age range. However, special equipment and expertise are required for the assessment of bone age. Hence, instead, surrogate measures of somatic maturation are generally used in nutrition surveys. WHO (1995) recommends the use of two maturational events for each sex to assist in interpreting anthropometric data during adolescence: one marker signaling the beginning of the adolescent growth spurt in each sex, and one indicating that the peak velocity for height and associated changes have passed. In girls, the indicator that can be used to signal that the adolescent growth spurt has begun is the start of breast development, which precedes peak height velocity by about 1y. The marker indicating that most of the adolescent growth spurt has been completed is the attainment of

menarche, which begins a little more than 1y after peak height velocity (Figure 9.1). In boys, the corresponding indicators signaling the beginning and completion of the adolescent growth spurt are adolescent changes in the penis, characterizing G3, followed by the attainment of adult voice, respectively (Figure 9.1).

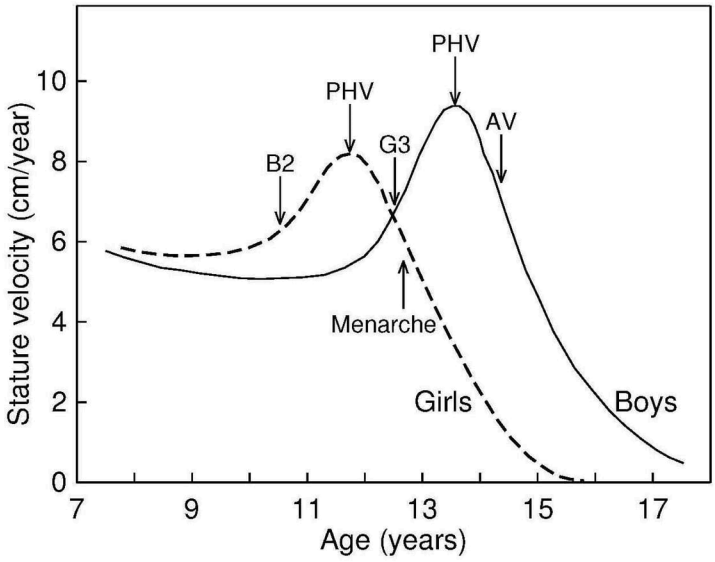


Figure 9.1. Approximate timing of recommended maturational events relative to peak height velocity (PHV) in boys and girls. B2 = Start of breast development. Identified by examination. Precedes PHV by about 1y. Indicates adolescent spurt has begun. Menarche. Determined by questioning. Menstruation usually begins a little more than 1y after PHV. Indicates that most of the adolescent spurt has been completed. G3 = Adolescent changes in the penis, characterizing G3. Identified by examination. Precedes PHV by about 1y. Indicates adolescent spurt has begun. AV= Attainment of adult voice. Determined by questioning. Usually attained about 1y after PHV. Indicates that most of the adolescent spurt has been completed. Redrawn from: WHO (1995).

When an assessment of somatic maturation cannot be obtained by physical examination and questioning, then a self-administered questionnaire containing drawings illustrating Tanner's stages of development of breasts and pubic hair for females, or pubic hair and male genitalia, may be used. Adolescents are requested to select the drawing closest to their stage of development, as described in Morris and Udry (1980).

9.4.2 Appropriate anthropometric reference data

In public health settings, appropriate anthropometric reference data facilitate international comparisons of anthropometric indices across populations and enable the proportion of individuals with abnormal indices to be determined relative to the reference population. Such comparisons enable the extent and severity of malnutrition in the study group to be estimated. In surveillance studies, reference data allow the evaluation of trends over time, as well as the effectiveness of intervention programs to be assessed. Reference data can also be used in clinical settings to monitor growth of individuals, detect abnormal changes in growth, and assess response to treatment (WHO, 1995).

The WHO recommends the use of the WHO Child Growth Standards for young children from birth to 5y for international use (WHO, 2006) in view of the small effect of ethnic and genetic differences on the growth of infants and young children compared with the environmental, nutritional, and socio-economic effects, some of which may persist across generations. The WHO Child Growth Standard was developed as a result of the technical and biological limitations identified with the earlier NCHS/WHO growth reference (Garza and de Onis, 1999). A prescriptive approach depicting physiological human growth under optimal conditions was used for the new Child Growth Standards so they represent how young children should grow, rather than as a “reference” describing how children do grow. To achieve this goal, a set of individual eligibility criteria were developed: term singleton infants with non-smoking mothers, a health status that did not constrain growth, and mothers who were willing to follow current WHO feeding recommendations. The design combined a longitudinal study from birth to 24mos with a cross-sectional study of children aged18–71mos based on pooled data from 6 participating countries (Brazil, Ghana, India, Norway, Oman, and the United States) (de Onis et al., 2004). WHO has developed a tool for the application of the WHO Child Growth Standards which includes

instructions on how to take the measurements, interpret growth indicators, investigate causes of growth problems, and how to counsel caregivers. An anthropometry training video is also available. For more details, see: WHO Child Growth Training Module.

For older children, the WHO growth reference data for school-age children and adolescents 5–19y should be used (de Onis et al., 2007). This is a reconstruction of the original 1977 National Centre for Health Statistics (NCHS) data set supplemented with data from the WHO Child Growth Standard. The statistical methodology used to construct this reference was the same as that used for the WHO Child Growth Standard.

A series of prescriptive standards for monitoring fetal, newborn growth, and gestational weight gain have also been developed for international use by the INTERGROWTH-21st project. This project adhered to the WHO recommendations for assessing human size and growth, and followed healthy pregnant women longitudinally from 9wks of fetal life to 2y (Papageorghiou et al., 2018; Ismail et al., 2016). Populations from urban areas of 8 countries in which maternal health care and nutritional needs were met (Brazil, China, India, Italy, Kenya, Oman, the UK and the USA) were involved to ensure universal multi-ethnic growth standards were generated that represent how fetuses should grow. Postnatal growth standards for preterm infants were also developed by this group (Villar et al., 2015).

Updated childhood growth charts have been prepared by the Center for Disease Control (CDC) for U.S children for two age groups: 0–36mos and2–20y. These CDC 2000 growth charts are based primarily on physical measurements taken during five nationally representative surveys conducted between 1963 and 1994, although some supplemental data were also used. When creating these revised growth charts, two data sets were excluded: growth data for very low birthweight infants (< 1500g) whose growth differs from that of normal birth-weight infants, and weight data for children > 6y who participated in the NHANES III survey. The latter data were excluded from both the revised weight and BMI growth charts because their inclusion shifted the upper percentile curves. Hence, the exclusion of these selected data resulted in a modified growth reference that is not a purely descriptive growth reference because it does not contain representative national data for all variables (Kuczmarski et al., 2000). A comparison of these CDC 2000 Growth Charts with the WHO Child Growth Standards is available in de Onis et al. (2007).

For body composition indices, use of local reference data are preferred because racial differences exist in both body proportions, and the amount and distribution of subcutaneous and intra-abdominal fat (Wagner and Heyward, 2000; He et al., 2002; Lim et al., 2019). In practice, however, only a few countries have local body composition reference data. In the absence of such local data, WHO recommend the use of the reference data for mid-upper-arm circumference (MUAC), triceps and subscapular skinfolds based on the data collected during the WHO MGRS on children age 0–5y.Electronic copies of the WHO tables and charts of percentiles and Z-scores for MUAC-for-age, triceps-for-age, and subscapular-for age by sex are available for children from age 3mos to 71mos and are included in WHO Child Growth Standards.

In clinical settings, abnormal changes in the rate of growth of a child can be detected much earlier when growth velocity charts, rather than distance growth charts, are used; see Chapter 13 for more details. Growth velocity charts are based on longitudinal studies during which the same child is measured serially, and the growth rate calculated for each interval. WHO has developed a set of growth velocity charts based on the WHO MGRS described earlier for international use; see de Onis et al. (2011) for more details.

Several different distance growth standards have been compiled, depending on the specific determinants of growth. For example, the new international postnatal growth standards should be used for preterm infants (Villar et al., 2015) in view of the difference in weight, length, and head circumference between preterm and full-term infants. Moreover, the time period over which these differences extend varies with the growth measurement. Differences are significant until 18 mos for head circumference, until 24mos for weight-for-age, and up to 3.5y for length/height-for-age.

Alternatively, tempo-conditional growth charts can be used for monitoring the growth of individual children during adolescence (Figure 9.2).

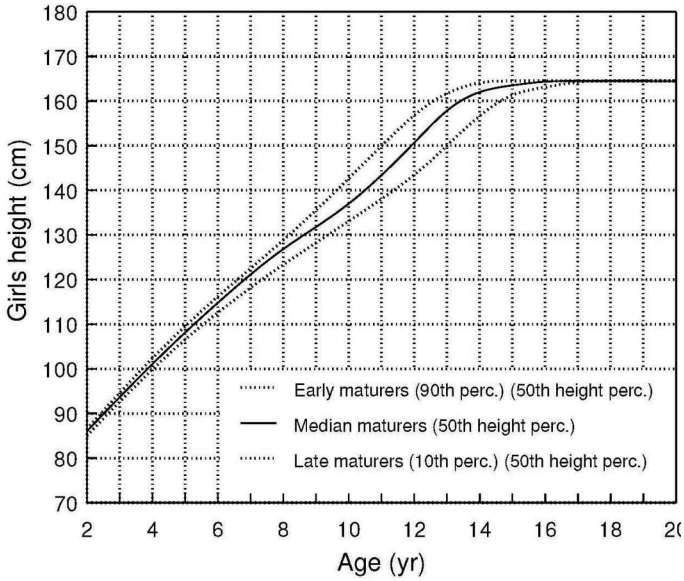


Figure 9.2. Diagrammatic height percentile chart for early-, median-, and late-maturing girls. Adapted from Tanner and Buckler (1997).

These growth charts are based on mixed cross-sectional and longitudinal data, and take into account differences in the timing of the adolescent growth spurt — termed the “phase difference effect” (Tanner & Buckler, 1997).

Parent-allowed-for growth reference data are available for children from 2–9y when nonfamilial short stature is of concern. Cole (2000) developed a novel parent-allowed-for-height chart that adjusts for mid-parent, single parent, or sibling height based on the UK90-height reference. Special growth charts have also been compiled for children with certain genetic disorders such as Down's syndrome or other developmental disorders in which growth patterns differ from the reference growth curves.

9.4.3 Methods of evaluating anthropometric indices

For studies of both individuals and populations, the anthropometric indices can be compared to the reference population using percentiles or Z-scores derived from the distribution of the anthropometric reference data. A percentile refers to the position of the measurement value in relation to all the measurements for the reference population, ranked in order of magnitude. A Z-score (or standard deviation score) measures the deviation of the value for an individual from the median value of the reference population, divided by the standard deviation of the reference, as shown below:

$$\text{Z-score or SD score} = \frac{(\text{ observed value }) - (\text{ median reference value })}{(\text{ standard deviation of reference population })}$$

In most industrialized countries, percentiles are used because no errors are introduced if the data have a skewed distribution. Weight-for-age, weight-for-height, and many circumferential and skinfold indices have skewed distributions.

Percentiles are not appropriate for use in low- and middle-income countries (LMICs) where many children may fall below the lowest percentile. In these settings Z-scores should be used because they can be calculated accurately beyond the limits of the original reference data. Hence, individuals with indices below the extreme percentiles of the reference data can then be classified accurately. For more discussion of percentiles and Z-scores, see Chapter 13. Both percentiles and Z-scores based on the WHO Child Growth Standards (0-5y) and the WHO Growth Reference for school-aged children and adolescents (5-19y) can be readily calculated in population studies using the WHO software program: WHO AnthroPlus (2009). Alternatively, for individuals in clinical settings, the percentile or Z-score range within which the measurement of an individual falls can be read from sex-specific charts or tables of the appropriate reference data, as shown in Figure 9.3.

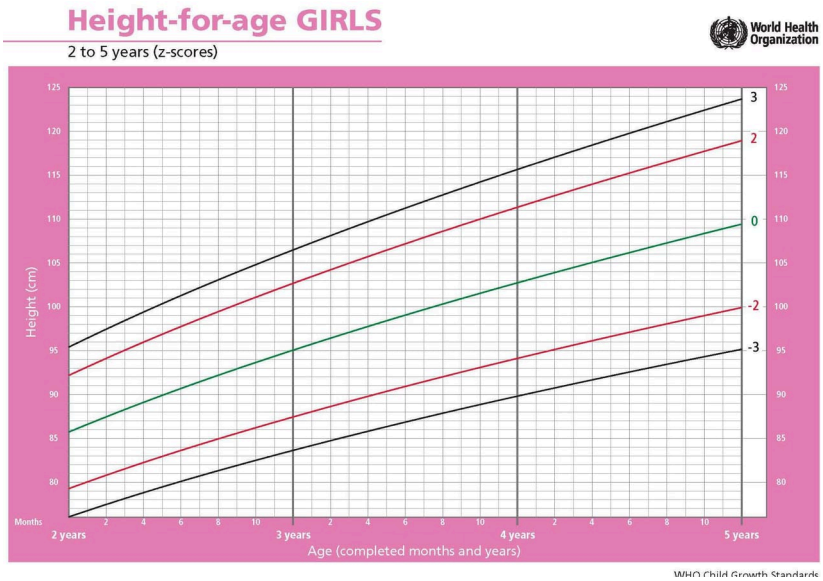


Figure 9.3. WHO Growth chart for girls aged 2&ndash5y for height-for-age. From: Training Course on Child Growth Assessment (WHO).

Three methods have been recommended by WHO to evaluate cross-sectional anthropometric data for use in public health; these are summarized in Box 9.4.

Box 9.4. Methods recommended by WHO to assess anthropometry

- Comparison of the frequency distribution of anthropometric indices with appropriate WHO growth reference data using Z-scores
- Summary statistics of the Z-scores: mean; median; SD; standard error (SE) with the 95% confidence interval (CI) for each growth indicator
- Calculation of number and proportion of individuals (as %)with anthropometric indices below or above a designated cutoff (95% CIs) for each age and sex group.

Figure 9.4 is an example of the first method summarized in Box 9.4. Here the frequency distribution of the Z-scores for height-for-age children from the Indian National Family Health Survey (2005-2006) are compared with the corresponding reference distribution of height-for-age Z-scores for the WHO Child Growth Standards. The figure highlights that nearly all the children surveyed were affected by some degree of linear growth retardation and would benefit from an intervention; this approach is termed a “population approach to targeting”.

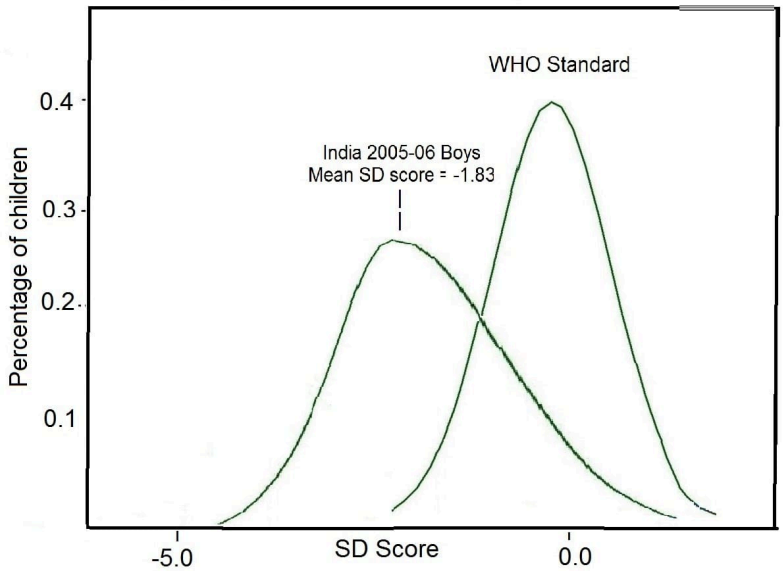


Figure 9.4 distribution of length/height-for-age Z-scores of children from the Indian National Family Health Survey 2005–2006. Modified from de Onis and Branca (2016)

Summary statistics can also be used when anthropometric indices are expressed as Z-scores in population studies. In the example given in Figure 9.4, the calculated mean height-for-age Z-score (-1.83) for the children was markedly lower than zero — the expected value for the reference distribution. This statistic alone indicates that the entire distribution has shifted downward, suggesting that most, if not all of the individuals, are affected by linear growth retardation, as was clear from the frequency distribution of the height-for-age Z-scores compared with the corresponding reference distribution depicted, in Figure 9.4.

Note, summary statistics cannot be calculated in this way for data from a population expressed in terms of percentiles which are often not normally distributed.

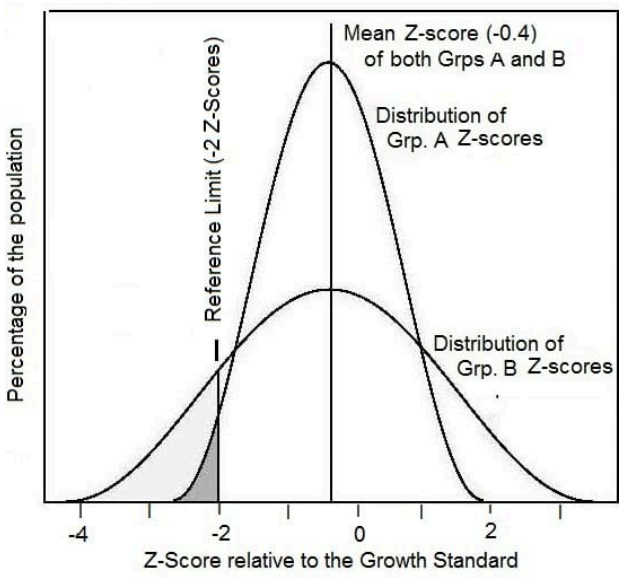


Figure 9.5 Two populations with the same mean Z-score, but different standard deviations.

Alternatively, calculation of the mean Z-score and SD can be used to compare directly different populations or the status of the same population at different times (Goldstein and Tanner, (1980). However, even if the populations have the same mean Z-score, their SDs may differ, with the population with the larger SD having a greater proportion below the reference limit or cutoff point, as shown in (Figure 9.5). Here, the growth reference is not being used for comparative purposes as shown in Figure 9.4.

The third method itemized in Box 9.4 involves calculating the percentage of individuals with anthropometric indices below or above predetermined reference limits or cutoff points. When used in this way, the anthropometric index and its associated reference limit or cutoff point are termed an "indicator" as described earlier. This approach is used to classify individuals as "at risk" to malnutrition and is used by governments and International Agencies (e.g., WHO and UNICEF) to generate prevalence estimates of malnutrition for information and comparisons across countries, as well as for advocacy. The approach is described in more detail below.

9.4.4 Classification systems

Classification systems are used in both clinical settings and in public health. All use at least one anthropometric measurement and one or more reference limit derived from appropriate reference data (i.e., indicator) to classify at risk individuals. Alternatively, cut-off points are used. In practice, classification schemes are not perfect, some misclassification will always occur so that some individuals identified as "at risk" to malnutrition will not be truly malnourished (false positives), and others classified as "not at risk" to malnutrition will in fact be malnourished (false negatives). Misclassification arises because there is always biological variation among individuals (and hence in the normal levels defined by the measurement) (Fraser, 2004); see Chapter 13 for more details.

Reference limits for anthropometric indices are derived from a reference distribution and can be expressed in terms of Z-scores or percentiles. In low income countries reference limits defined by Z-scores are frequently applied, with scores below –2 or above +2 Z-scores of the WHO Child Growth Standard or WHO Growth Reference used to designate individuals with either unusually low or unusually high anthropometric indices (WHO, 1995). This approach is used because statistically 95% of the international reference population fall within the central range assumed to be "healthy". Therore, theoretically, the proportion of children with a Z-score less than –2 or greater than +2 in a study population should be ≈ 2.3%. Clearly, if the proportion in the study population with such low or high Z-scores is significantly greater than this, then the study population is seriously affected. The WHO uses the below –2 Z-scores of the WHO reference median for weight-for-age, length/height-for-age, or weight-for-height to classify children as underweight, stunted, or wasted, respectively.

In industrialized countries, the percentiles commonly used for designating individuals as "at risk" to malnutrition are either below the 3rd or 5th and above the 97th or 95th percentiles. The limits chosen depend on the reference data used: see Chapter 13 for more details.

There is nothing immutable about a reference limit at –2 Z-score, despite their use by countries and agencies to generate prevalence estimates, most notably for stunting and wasting. As a consequence, attempts have been made to establish "cutoffs" for some anthropometric indicators to improve the ability to discriminate between children who are malnourished and those who are "healthy". These cutoffs have been established by a review of the anthropometric characteristics of individuals with either clinically moderate or severe malnutrition or who subsequently die. However, many other characteristics of individuals such as age, sex, life-stage, race/ethnicity, genetics, and morbidity or nutritional status may affect the relationship under study (Hondru et al., 2019; Yaghootkar et al., 2020; Wright et al., 2021). Hence, in practice, defining cutoffs is difficult because the relationship between the indices and the biological factors cannot be generalized from one region to another (WHO Expert Consultation, 2004).. Consequently, in some studies universal cutoff points are used, whereas in others the methods used to identify the cutoff points applied are not always well documented.

As an example in a study of low BMI and morbidity in Pakistan, a reported a cutoff of < 18.5 was associated with higher morbidity, whereas in Calcutta it was < 16.0 (Campbell and Ulijaszek, 1994; Kennedy and Garcia, 1994).

Cut-offs for BMI, waist circumference, and waist-hip ratio associated with risk of cardiovascular disease and type 2 diabetes have

also been extensively investigated among different ethnic groups (Ding et al., 2020; WHO Expert Consultation, 2004; Lear et al., 2010). Some expert groups have defined lower waist circumference cutoffs for adults of Asian descent compared to Europeans (IDF, 2006).

Receiver operating characteristic (ROC) curves are often used to determine cutoff points. This is a graphical method of comparing indices and portraying the trade-offs that occur in the sensitivity and specificity of a measurement or index when the cutoffs are altered. To use this approach, a spectrum of cutoffs over the observed range of the indicator results is used, and the sensitivity and specificity for each cutoff calculated. Next, the sensitivity (or true-positive) rate is plotted on the vertical axis against the true negative rate (1–specificity) on the horizontal axis for each cutoff point, as shown in (Figure 9.6).

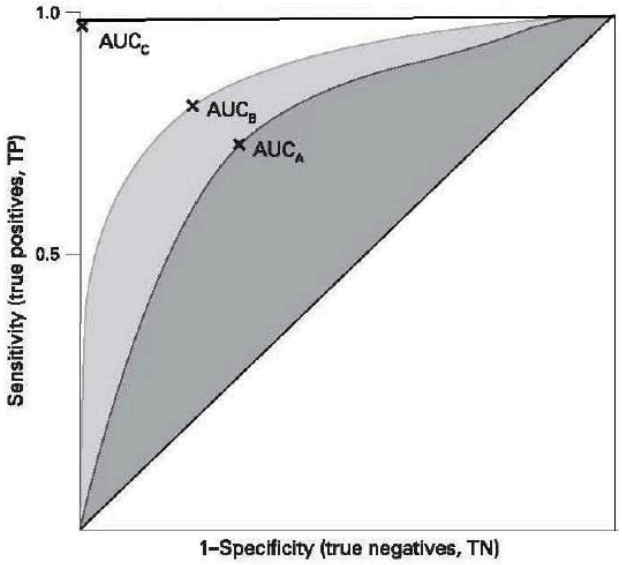


Figure 9.6. Receiver-operating characteristic curves. Three plots and their respective areas under the curve (AUC) are given. The diagnostic accuracy of marker C (white area) is better than that of B and A, as the AUC of C > B > A. X = point of best cut-off for the biomarker. From: Søreide, 2009, with permission of the BMJ Publishing Group Ltd.

The closer the curve follows the left-hand of the ROC space, the more accurate is the cutoff under study in distinguishing the health or nutritional status condition under investigation from optimal status. The optimal ROC curve is the line connecting the points highest and farthest to the left of the upper corner. The closer the curve comes to the 45° diagonal of the ROC space, the less accurate the indicator cutoffs (Søreide, 2009). Most statistical programs (e.g., SPSS) provide ROC curve analysis. Details of alternative statistical methods for selecting the best cutoff point are given in Brownie et al. (1986).

The choice of the cutoff may vary depending on the circumstances. When resources are scarce, a low cutoff point may be selected. As a consequence, the sensitivity decreases, which mean that more truly malnourished children are missed. However, at the same time, the specificity increases, which means that fewer well-nourished children are misdiagnosed as malnourished. Conversely, when resources are generous, the cutoff can be high, because it does not matter if some children receive treatment when they do not need it. The inverse relationship between sensitivity and specificity and relative risk of mortality associated with various values for MUAC in children6–36mos in rural Bangladesh is shown in Table 9.4. More details of the inter-relationships between these variables is given in Chapter 1.

Table 9.4. Sensitivity, specificity, and relative risk of death associated with various values for mid-upper-arm circumference in children 6–36mos in rural Bangladesh. Data from Briend et al.(1987).

Arm circumference (mm)	Sensitivity (%)	Specificity (%)	Relative Risk of death
≤ 100	42	99	48
100–110	56	94	20
110–120	77	77	11

Arm circumference (mm)	Sensitivity (%)	Specificity (%)	Relative Risk of death
120–130	90	40	6

Unfortunately, because sensitivity and specificity data for the anthropometric indices selected are usually not known for the population under study, the data required to plot ROC curves are often obtained elsewhere, even though the values may not be appropriate for the population under study because of the many factors known to influence cutoff values, as noted earlier; see Chapter 13 for more details. Ultimately the choice of an index and an associated cutoff point (i.e., an indicator) depends on the resources available and the purpose for which it is being used. The latter can range from screening for disease or monitoring to detect changes in prevalence of malnutrition etc. For more discussion on the selection of anthropometric indicators, see Brownie et al. (1986) and WHO (1995).

Screening tools to identify those who are malnourished in clinical or public health settings can be based on single or multiple measurements and associated reference limits or cutoff points. An example of a screening tool used in clinical settings is the Malnutrition Universal Screening Tool (MUST), widely used to identify adults who are at risk of undernutrition or obesity (See Chapter 27). MUST is based on height and weight measurements from which both a BMI score (from 3 cutoffs) assessed via a chart and weight loss in previous 3–6mos, can be derived, followed by establishing acute disease effect and score. From the sum of the three scores, the overall risk of malnutrition is calculated, with management guidelines provided according to the level of risk (low; medium; high). More details are available: Malnutrition Universal Screening Tool .

In public health emergencies, a screening tool based on a single measurement (i.e., MUAC) and associated cutoff (i.e., < 115mm) is often used to identify severe acute malnutrition (SAM) in children 6–60mos. This MUAC cutoff was chosen because children with a MUAC < 115mm were observed to have a highly elevated risk of death compared to those with a MUAC > 115mm (Myatt et al., 2006).

For defining overweight and obesity in children and adolescents, WHO recommends the use of BMI-for-age Z-scores and reference limits based on Z-scores. For children (0–5y), a Z-score for BMI-for-age above +1 is described as being “at risk of overweight”, above +2 as “overweight”, and above +3 as “obese” based on the WHO Child Growth Standard. For children 5–19y, BMI-for-age Z-scores above +1 and above +2 based on the WHO 2007 growth reference data are recommended (de Onis & Lobstein, 2010). To classify overweight and obesity in adults, however, WHO recommends a graded classification scheme, as shown in (Table 9.5).

Table 9.5. WHO classification of obesity in adults according to body mass index (BMI). From: WHO (2000).

Classification	BMI (kg/m^2)	Risk of comorbidities
Underweight	< 18.50	Low (but risk of clinical problems is increased)
Normal range	18.50–24.99	Average
Overweight	**≥ 25.00**	
Pre-obese	25.00–29.99	Increased
Obese class I	30.00–34.99	Moderate
Obese class II	35.00–39.99	Severe
Obese class III	≥ 40.00	Very severe

Increasingly, it is recognized that in low-income countries, multiple anthropometric deficits may occur simultaneously in children and amplify their risk of morbidity and mortality. Consequently, a composite index of anthropometric failure (CIAF) has been developed, and is described in Chapter 13.

In public health, screening tools are also used to map countries according to levels of severity of malnutrition (UNICEF/WHO/World Bank, 2021) in order to identify priority countries. Five prevalence thresholds (as %) based wasting (i.e., WHZ < –2), over-

weight (BMIZ > +2), and stunting (i.e., HAZ < –2) have been developed by WHO and UNICEF; these are depicted in (Table 9.6). The fifth threshold labelled “very low” and of no public health concern was included across all three indicators to reflect the expected prevalence of 2.3% (rounded to 2.5%) below/above 2 SDs from the median of the WHO Child Growth Standard.

Table 9.6. Prevalence thresholds, corresponding labels, and the number of countries (n) in different prevalence threshold categories for wasting, overweight and stunting in children under 5y using the “novel approach”. From: de Onis et al. (2018).

Wasting			Overweight			Stunting		
Prevalence thresholds (%)	**Labels**	**(n)**	**Prevalence thresholds (%)**	**Labels**	**(n)**	**Prevalence thresholds (%)**	**Labels**	**(n)**
< 2·5	Very low	36	< 2·5	Very low	18	< 2·5	Very low	4
2·5 – < 5	Low	33	2·5 – < 5	Low	33	2·5 – < 10	Low	26
5 – < 10	Medium	39	5 – < 10	Medium	50	10 – < 20	Medium	30
10 – < 15	High	14	10 – < 15	High	18	20 – < 30	High	30
≥ 15	Very high	10	≥ 15	Very high	9	≥ 30	Very high	44

The number of countries in different threshold categories for wasting, overweight, and stunting, also shown in (Table 9.6), is based on data from 134 countries. Comparison of the prevalence estimates for each anthropometric indicator can trigger countries to identify the most appropriate intervention program to achieve “low” or “very low” prevalence thresholds.

Details of the techniques used to measure body size and body composition, together with the indices derived from these measurements, are discussed in Chapters 10 and 11, whereas Chapter 13 discusses methods used for evaluation of anthropometric indices and their application at the individual and population level.

This page titled 8.4: Interpretation and evaluation of anthropometric data is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

9: BODY SIZE (CHAPTER 10)

Abstract

Body size can be assessed from measurements of stature (height or length), weight, and head circumference using standardized procedures described here. When height cannot be measured, published equations based on knee height, lower leg length, arm span or demi-span are used. To interpret the measurements, indices such as weight-for-age, body-mass-index-for-age, stature-for age, weight-for stature, and body mass index (BMI, weight / height^2) are constructed.

To evaluate the anthropometric indices, they can be compared with appropriate growth reference data using Z-scores or percentiles. For international use, the WHO prescriptive Child Growth Standard for children (0–5y) and the WHO growth reference for school-age children and adolescents (5–19y) should be used. Statistically determined reference limits or cutoffs can also be applied to generate "anthropometric indicators" such as stunting (stature-for-age < –2 Z-score), wasting (weight-for-stature < –2 Z-score), overweight (weight-for-stature > +2 Z-score), and thinness, overweight, and obesity based on BMI-for-age < –2, > +2, and > +3 Z-scores, respectively. Both stature-for-age at < – Z-score and weight-for-stature at < –2 Z-score are recommended by WHO because together they can distinguish between stunting and wasting. To assess changes in stature over time, however, use of stature-for-age difference rather than stature-for-age Z-score is now preferred.

To assess and compare the severity of malnutrition across countries, five prevalence thresholds for wasting, overweight, and stunting can be used. Recognition of an elevated mortality risk for children suffering from multiple anthropometric deficits simultaneously has led to the development of the Composite Index of Anthropometric Growth Failure (CIAF). This is used to characterize the overall burden of under- and -over-nutrition in children age < 5y so they can be identified and prioritized for intervention.

Changes in body weight over time are used in both clinical and public health settings. In clinical settings lack of weight gain and weight loss are used identify pediatric undernutrition, whereas weekly weight gain is used to monitor response of children with severe acute malnutrition (SAM) to feeding programs. In public health, changes in body weight over time are often monitored because of their increased risk of mortality. New U.S guidelines for gestational weight gain are now available which have been adopted by WHO for low-income populations until specific country cutoffs are available.

BMI is currently the indicator of choice for defining underweight, overweight, and obesity in adults and children. However, the limitations of BMI alone has led to the use of both BMI and waist circumference because the distribution of abdominal visceral fat has an even greater cardio-metabolic risk than excess body fat *per se*. The BMI cutoffs used to classify adults as underweight (BMI < 18.5kg/m^2), overweight (BMI > 25kg/m^2), and obese (BMI > 30kg/m^2) are based on adverse health risks and sometimes risk of mortality. In adults, WHO recommends the use of universal BMI cutoffs for overweight (BMI > 25kg/m^2) and obesity (BMI > 30kg/m^2); three classes of obesity are also defined. This classification has been adopted by several countries, including Canada and the United States, with the highest class indicative of severe obesity (i.e., BMI >

40kg/m^2). Cutoffs to define grades of low BMI values indicative of underweight in adults are also recommended together with a classification to identify populations with a public health problem. Increasingly, low- and middle-income countries are impacted by both under- and over-nutrition, a condition termed the “double burden of malnutrition”.

In children, no simple cutoff to define thinness, overweight, or obesity can be used because BMI has a characteristic curvilinear shape with age. Instead, countries have compiled their own ways of defining thinness, overweight, and obesity in children. Three classification schemes are described with cutoffs defined by BMI-for-age Z-scores or percentiles based on international or national growth reference data. They include those set by the International Obesity Task Force (IOTF), the new WHO classification based on their international growth standard and growth reference, and the U.S classification based on the 2000 CDC growth charts. Clearly, there is an urgent need for consensus on BMI cutoffs for thinness, overweight, and obesity in childhood that are defined by adverse health risks. Only in this way can valid international comparisons across countries on the prevalence of thinness, overweight, and obesity in childhood be made.

9.1: Anthropometric assessment of body size (10.0)

9.2: Measurements of body size (10.1)

9.3: Growth indices, indicators, and recommended growth reference data (10.2)

9.4: Body mass index in adults (10.3)

9.5: BMI in children and adolescents (10.4)

This page titled 9: Body size (Chapter 10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Anthropometric assessment of body size (10.0)

The most widely used anthropometric measurements of body size are those of stature (height or length) and body weight. These measurements can be made quickly and easily and, with care and training, accurately. Head circumference measurements are often taken in association with stature. Details of standardized procedures for these measurements of body size are summarized below and are given in detail in(Lohman et al., 1988).

Indices such as head-circumference-for-age, weight-for-age, weight-for-stature, and stature-for-age and the ratio of weight to stature are derived from these measurements. Of these, stature-for-age and weight-for-stature have been recommended by the World Health Organization (WHO, 1995a) for use in low-income countries. In combination, they can distinguish between stunting and wasting. There is also recognition that individual children may be simultaneously wasted and stunted, which has prompted interest in identifying the combination of anthropometric deficits with the greatest risk of mortality (McDonald et al., 2013).

In epidemiological studies body mass index (BMI) (wt/ht^2, as kg/m^2) is used to define thinness, overweight, and obesity. Increasingly, BMI is accompanied by waist circumference to provide an assessment of abdominal fat in children, adolescents, and adults. In hospital settings, anthropometric indices of body size are used primarily to identify under- or over-nutrition and obesity, and to monitor changes after a nutrition intervention.

This page titled 9.1: Anthropometric assessment of body size (10.0) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

9.2: Measurements of body size (10.1)

Of the various measurements of body size, head circumference is important because it is closely related to brain size. It is often used with other measurements to detect the pathological conditions associated with either an unusually large (macrocephalic) or small (microcephalic) head.

Recumbent length is measured in infants and children < 2y. Height is measured in older children and adults. The interpretation of length (and weight) at birth and during later infancy requires a valid and precise estimate of gestational age: making such an estimate is often difficult in many low-income countries. To assess growth over short time periods, lower-leg length in infants and children can be measured. In adults, knee-height measurements are used to estimate height in those persons with severe spinal curvature and in those who are unable to stand. Alternatively, arm span or demi-arm span can be measured when actual height cannot be performed.

Weight in infants and young children can be measured using a suspended scale and a weighing sling, or for greater precision (within 10g), a pediatric scale. For older children and adults, portable electronic digital scales with a taring capacity are recommended. Elbow breadth is used as a measure of frame size, which is relatively independent of adiposity and age (Frisancho, 1990); it should be measured with flat-bladed sliding calipers.

For details on the measurement techniques and standardization protocols used in the World Health Organization (WHO) Multicenter Growth Reference Study (MGRS), the reader is referred to: the WHO Anthropometry training course available from WHO (Training Course).

10.1.1 Head circumference

For the measurement of head circumference, a narrow, flexible, non-stretch tape made of fiberglass or steel (range 0–200cm, calibrated to 1mm, and about 0.7cm wide) should be used. Any head-bands or hair-pins should be removed and braids undone for the measurement.

An infant or child below the age of two years should be held on the mother's lap, whereas older children can stand with the left side facing the measurer, with arms relaxed and legs apart. The child must look straight ahead so the line of vision is perpendicular to the body and the Frankfurt plane — that is, an imaginary plane which passes through the external auditory meatus (the small flap of skin on the forward edge of the ear) and over the top of the lower bone of the eye socket immediately under the eye — is horizontal. The tape is placed just above the supraorbital ridges (i.e., slightly above the eyebrows) covering the most prominent part of the frontal bulge and over the part of the occiput that gives the maximum circumference (i.e., fullest protuberance of the skull at the back of the head) Figure 10.1. Care must be taken to ensure that the tape is at the same level on each side of the head and pulled tightly to compress the hair and skin. Measurements are recorded to the last completed 1mm (de Onis et al., 2004).

Where possible, serial measurements of head circumference should be incorporated into routine well-child care, to establish if head circumference is within normal limits, too large (megacephaly), or too small (microcephaly) (Harris, 2015). For international use, the WHO Child Growth Standards that include head circumference measurements are recommended (Growth Standards).

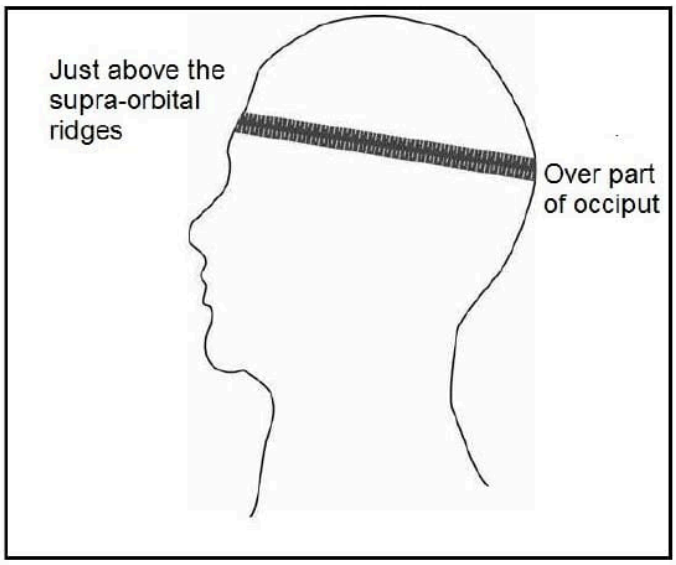


Figure 10.1. Measurement of head circumference.

10.1.2 Gestational age

The assessment of gestational age is necessary for the interpretation of any size-for-age measurement of infants and for following the neurodevelopmental progress of preterm infants. It is also essential for the management of pregnancy and treatment of the newborn.

Several strategies are available for estimating gestational age. Prenatal measures of gestational age include calculating the number of completed weeks since the beginning of the last menstrual period, prenatal ultrasonography, and clinical methods. Of these, the definition of gestational age on the basis of the last menstrual period (LMP) is most frequently used in low-income countries, but it is associated with several problems: errors may occur because of irregular menses, bleeding early in pregnancy, and incorrect recall by mothers. Macaulay et al. (2019) concluded that use of LMP for gestational age dating during early pregnancy was not sensitive in identifying late- and post-term pregnancies based on a comparison of gestational age estimates from LMP and ultrasonography.

Prenatal ultrasonography during the first or second trimester is considered by many to be the gold standard for assessment of gestational age. Estimates are based on different ultrasonic measures of fetal size, with crown-rump length considered the best single parameter for the first trimester, whereas for a second- or third-trimester scan, a combination of multiple biometric parameters such as the biparietal diameter, head circumference, femoral length, and abdominal circumference are recommended (Butt & Lim, 2019). Measurements are most accurate when made early in gestation.

Fetal growth charts based on ultrasound biometric measurements from multi-national prospective observational longitudinal studies have been developed by WHO (Kiserud et al., 2017) and by the INTERGROWTH-21st project (Papageorghiou et al., 2014). These fetal growth charts are based on serial ultrasound examinations during pregnancy in which women with obstetric conditions that may influence growth were excluded. The standards of the INTERGROWTH-21st project are based on populations from urban centers in eight countries in which maternal health care and the nutritional needs of women were met (Brazil, China, India, Italy, Kenya, Oman, the UK and the USA). The objective was to generate universal multi-ethnic growth standards that represent how fetuses should grow (i.e., the standards are prescriptive) when nutritional, environmental, and health constraints on growth are minimal. Fetal growth charts for estimated fetal weight and common ultrasound biometric measurements at 14–42wks of gestation are presented (Papageorghiou et al., 2014).

Unfortunately, ultrasonography is not universally available, especially in low-income countries, and the quality of both the equipment used and the technical training varies. Instead, clinical methods of prenatal assessment such as measurement of symphysis-fundal height (in cm) are often used, which should correlate with the week of gestation after 20wks. International standards for symphysis-fundal height based on serial measurements from the fetal growth longitudinal study of the INTERGROWTH-21st Project are also available (Papageorghiou et al., 2016). Alternative clinical methods sometimes used include the

auscultation of fetal heart tones (audible at 11–12wks), and the recording of fetal movements (Alexander & Allen, 1996; (Butt & Lim, 2019).

Several scoring systems, based on external and neurological criteria, have been developed to estimate maturity — and thus gestational age — postnatally. The scoring systems initially devised by Dubowitz et al. (1970) and later modified by Ballard et al. (1979) have been widely adopted. Both methods appear to have limited accuracy at the extremes of gestation. Some (Tergestina et al., 2021) but not all (Stevenson et al., 2021) investigators advocate foot length, measured postnatally (< 48h) with calipers, as useful in resource-limited settings. To date, the measurement of gestational age postnatally where ultrasound is not possible continues to be problematic (Stevenson et al., 2021).

10.1.3 Recumbent length

For infants and children < 2y (i.e., < 85cm), recumbent length is measured, preferably using an infantometer with a range of 30–110cm, and preferably equipped with an electronic digital reader. Recumbent length should be recorded to the nearest millimeter, or even more precisely (i.e., 0.1mm) when possible. Wooden or perspex length measuring boards (Figure 10.2) can also be used, although they are rarely fitted with digital counters so are less reliable. Note that recumbent length for a child of about 2y is approximately 5mm greater on average than standing height for the same child (Haschke & van't Hof, 2000).

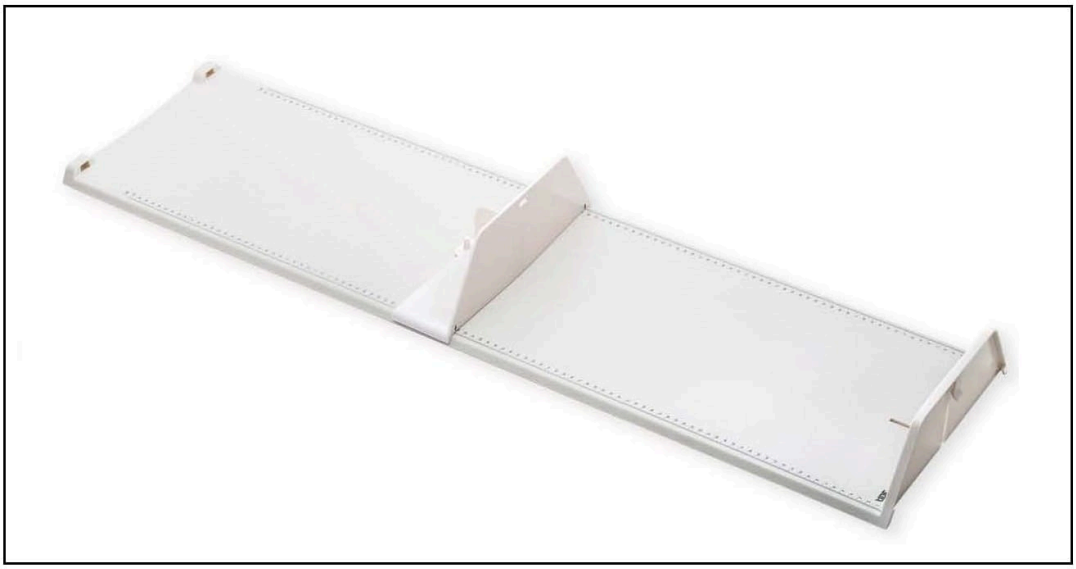


Figure 10.2. Device for the measurement of recumbent length. Available from: UNICEF.ORG.

Two examiners are required to correctly position the child and ensure accurate and reliable measurements of length. Prior to the measurement, braids must be undone and any hair ornaments removed. Diapers must also be removed so that both of the child's legs can be outstretched in the correct position. The subject is placed face upward, with the head toward the fixed end of the board and the body parallel to the board's axis. The shoulders should rest against the surface of the board. One examiner applies gentle traction to bring the crown of the child's head into contact with the fixed headboard and positions the head so that the Frankfurt plane is vertical (Figure 10.3).

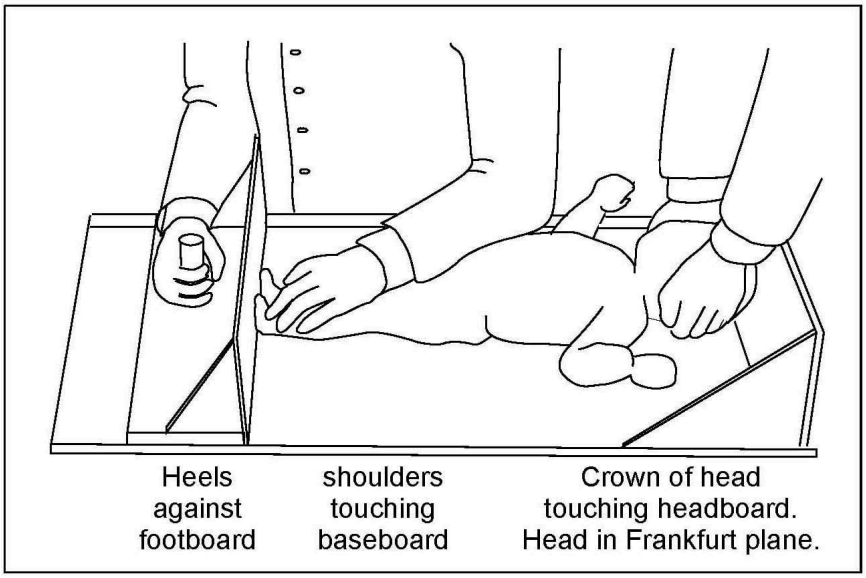


Figure 10.3. Measurement of recumbent length. Redrawn from Robbins GE, Trowbridge FL, in: Nutrition Assessment: A Comprehensive Guide for Planning Intervention by M.D. Simko, C. Cowell, and J.A. Gilbride (eds.), p.75, Aspen Publishers, Inc., 1984.

The second examiner holds the subject's feet, without shoes, toes pointing directly upward, and keeping the subject's knees straight, brings the movable footboard to rest firmly against the heels. To ensure the soles of the feet are flat on the footboard, the examiner should run the tip of their finger down the inside of the child's foot. The reading is taken to the nearest millimeter. If the subject is restless, only the left leg should be positioned for the measurement. To encourage and comfort the child during the measurement, the parent or guardian should stand between the examiner and recorder so that they can make eye contact and talk to the child during the measurement procedure.

10.1.4 Standing height

Children aged ≥ 2y and adults should be measured in the standing position, if possible (WHO, 1995a), using a free-standing stadiometer (range 65–200cm) with a fixed vertical backboard and an adjustable head piece, preferably equipped with a digital reader capable of measuring stature to 0.1mm.

Clothing should be minimal when measuring height so that posture can be clearly seen. Shoes and socks should not be worn, hair ornaments should be removed, and braids undone. The subject is asked to stand straight with the head in the Frankfurt plane (Figure 10.4), knees straight, with heels together and toes apart pointing slightly outward at a 60° angle. The back of the head, shoulder blades, buttocks, and heels must be in contact with the vertical backboard of the stadiometer. Arms should be hanging loosely at the sides with palms facing the thighs. For younger subjects, it may be necessary to hold the heels to ensure they do not leave the ground.

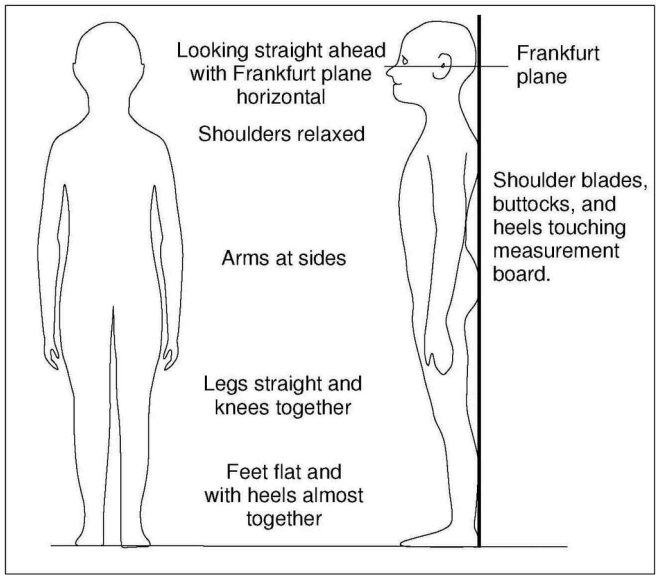


Figure 10.4. Positioning of subject for height measurement. The horizontal line is the Frankfurt plane, which should be in a horizontal position when height is measured. Reproduced from Robbins GE, Trowbridge FL, in: Nutrition Assessment: A Comprehensive Guide for Planning Intervention by M.D. Simko, C. Cowell, and J.A. Gilbride (eds.), p.77, Aspen Publishers, Inc., 1984.

Some investigators recommend applying gentle upward pressure to the mastoid processes to stretch the spine and minimize effects produced by diurnal variation(Tanner et al., 1966). Subjects are asked to take a deep breath and stand tall to aid the straightening of the spine. Shoulders should be relaxed. The movable headboard is then gently lowered until it touches the crown of the head. The height measurement is taken at maximum inspiration, with the examiner's eyes level with the headboard to avoid parallax errors. Height is recorded to the nearest millimeter, or even more precisely with more modern digital equipment.

The time at which the measurement is made should be recorded; diurnal variations in height occur due to compression of the spine as the day progresses. Consequently, for cross-sectional and longitudinal studies, heights should be measured at the same time of day, preferably in the afternoon. In cases where large amounts of adipose tissue prevent the heels, buttocks, and shoulders from simultaneously touching the wall, subjects should simply be asked to stand erect with the head in the Frankfurt plane.

In the field, vertical surfaces are not always available. In such circumstances, modified tape measures such as the Microtoise, which measure up to 2m, can be used. To use the Microtoise, it is first placed on the floor, after which the tape is pulled out to its fullest extent and released, and the end is fixed with a nail to a door or doorway. The subject is then instructed to stand erect directly below the point of attachment. An anthropometrist should position the subject's head correctly in the Frankfurt plane, as described in Section 10.1.1, before the tape is lowered by a second person until the head-bar touches the crown of the head and compresses the hair. A direct reading of height to the nearest millimeter may then be obtained.

Recumbent length for a child of about 2y is approximately 5mm greater on average than standing height for the same child (Haschke & van't Hof, 2000). Hence, if standing height rather than recumbent length is measured, 5mm must be added to the standing height value when recumbent length reference data are used. In general self-reported heights tend to produce slightly higher estimates of measured height, the magnitude of the discrepancy varying depending on age, race/ethnicity (Ekström et al., 2015; Hodge et al., 2020).

When measuring recumbent length or standing height, attempts should be made to minimize measurement errors. In longitudinal studies involving sequential measurements on the same group of individuals, it is preferable whenever possible, to have one person carrying out the measurements throughout the study to eliminate between-examiner errors. This is especially critical when growth velocity is calculated; growth increments are generally small and are associated with two error terms, one on each measurement occasion. Recommendations of the minimal intervals necessary to provide reliable data on growth increments during infancy and early childhood (de Onis et al., 2004) and adolescence are available. In the WHO Multicentre growth reference study (MGRS) for example, length was measured every two weeks from aged 2–6 weeks, monthly for ages 2–12 months, and every two months for toddlers aged 14–24 months. Increments measured over 6 months are the minimum interval that can be used to provide reliable data during adolescence. For shorter intervals, the combined errors may be too large in relation to the expected

mean increment. Length or height velocity is usually expressed as cm/y.

In large regional surveys, it is often necessary to use several well-trained anthropometrists. In such circumstances, the anthropometrists should be rotated among the subjects to reduce the effect of measurement bias of the individual examiners. Further, regular standardization sessions with the assessment of both within- and between-examiner reliability throughout the data collection should be conducted to maintain the quality of the measurements, and identify and correct systematic errors in the measurements; details of the procedures used in the WHO MGRS are given in de Onis et al. (2004). The WHO MGRS recommends that the maximum allowable difference in length for acceptable precision between measurements by two examiners is 7.0mm. This figure is based on the technical error of the measurement (TEM) obtained in the initial standardization session conducted at the Brazil site (de Onis et al. 2004). Details of the measurement techniques and standardization protocols for both recumbent length and stature are also available in an anthropometric training video prepared for the WHO MGRS and available on request from de Onis et al. (2004). Statistical methods exist for removing anthropometric measurement errors from cross-sectional anthropometric data; details are given in Ulijaszek and Lourie (1994).

10.1.5 Knee height in children aged over 3y

The measurement of knee height in children is termed “knemometry”. The measurement is taken from the distance between the heel and knee of the right leg (i.e. the lower leg length) when the child is sitting. At least three and preferably six measurements are needed for an accurate determination of the lower leg length in a child (Hermanussen et al., 1988). A training period of several weeks is required prior to routine measuring of lower leg length in children > 3y (Ahmed et al., 1995).

The main application of knemometry is for physically disabled children and in certain pediatric units specializing in growth disorders. Knemometry can also be used to assess the effect of therapeutic interventions (e.g., steroid therapy etc) on short-term growth (Gradman & Wolthers, 2010; Battin et al., 2012). By measuring knee height, growth increments in children are said to be detected more readily and over a shorter time frame than by conventional height measurements. Moreover, knee height measurements can be made with greater precision.

Several factors other than growth influence the measurement of lower leg length, and hence must be controlled. For example, as diurnal variations influence the measurement, it is preferable for all measurements to be performed during the afternoon, and by one trained operator. Before the measurements, children should avoid vigorous physical activity for at least 2h, and instead, stand or walk slowly for 5–10min.

To measure lower leg length in children > 3y, the children must be able to sit quietly and co-operate. The first knemometer for measuring children > 3y was developed by Valk in 1971, and modified in 1983 to improve its accuracy. Portable devices, termed a knee height measuring devices (KHMDs), are now also available for measuring changes in lower leg length in children > 3y (Cronk et al., 1989). Extremely low TEMs have been reported for the original Valk knemometer (0.09 and 0.16) (Hermanussen et al., 1988), whereas the newer, less costly portable knemometers, have a slightly poorer performance (Cronk et al., 1989). For children < 3y, a mini-knemometer can be used to measure lower leg length (Section 10.1.6).

Measurement using a knemometer

(Figure 10.5) depicts the technique of lower leg length measurement using a modified Valk knemometer. To take the measurement, the child is asked to sit on the chair. The sitting height (A) and the chair back (B) and chair position can be adjusted. The foot of the child is placed on the foot rest (C) and the angle and the length of the foot are recorded. The sitting position of the child is standardized by recording the following:

- sitting height of the child
- distance between the chair and the measuring board
- individual sitting position of the child (i.e., distance between the lateral condyle (X) and the back of the chair (B).

Next, the measuring board (D) is lowered onto the child's knee, and then the chair on which the child is sitting is gently moved forward and backward. The knee is moved passively under the measuring board (D) and the foot plate (C), while the length is displayed continuously on an electronic display. A counterweight (E) ensures a constant pressure of the measuring board of about 200g.

The actual lower leg length is defined as the maximum distance that can be reached during the movements of the child's knee. The minimum distance that can be discriminated by this instrument is 0.1mm. Children should be lightly dressed when these measurements are taken.

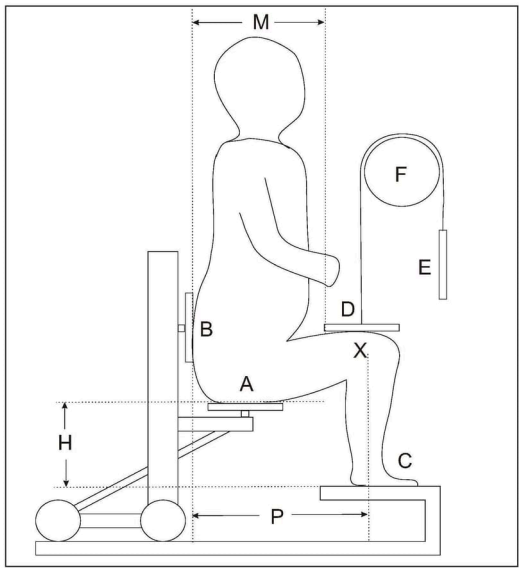


Figure 10.5. Diagram of the knemometer. Movable chair with adjustable height (A) and back (B); foot plate (C); Measuring board (D); Counterweight (E); Visual electronic display (F). The coordinates of the child's sitting position are defined by: the sitting height (H); the distance between the chair and the measuring board (M); and the distance between the lateral condyle (marked "X") and the back of the chair (P). Redrawn from Hermanussen et al., Annals of Human Biology 15: 1–16, 1988.

Portable knee height measurement device (KHMD)

The portable instrument is also designed to measure short-term growth of the lower leg. This device is less costly and easier to use than the knemometer. Again, the measurements are taken while the child is sitting. The chair used with this device should have a seat height of 33cm and a seat length of 26cm.

The KHMD consists of a standard industrial electronic height gauge (precision of 0.01mm) to measure the distance between the knee and foot plate. A plexiglass knee plate is mounted on the height gauge. A dial indicator is attached to the height gauge, so accurate and repeatable force can be applied to the knee during each measurement. The reading on the dial indicator corresponds to the torque flexion of the knee plate during the measurement.

The height gauge with the knee plate attached is mounted on a base comprised of a foot plate, wheels, and a handgrip to move the device. Both heel and toe plates are mounted on the foot plate to ensure that the position of the foot is reproducible for each measurement. The knee is also fixed in position during the measurement by an adjustable knee brace. The knee brace position is recorded by the operator from a numerical scale attached to the upright support. This allows the operator to position the subject reproducibly at each measurement session.

The position of the chair relative to the KHMD must also be fixed for each subject at each measurement session. This is achieved by placing the chair on a plexiglass grid with numbered and lettered squares. This marked grid, together with the braces for the knee, foot, and toe, ensure that the position of the subject at each measurement session can be replicated (Cronk et al., 1989).

For the measurement, subjects should sit with buttocks and the small of the back firmly in contact with the chair back, with the foot on the foot plate flush with the heel brace. The top of the knee brace is then adjusted until it has contact with the nonboney space between the patella and the tibial tuberosity. The chair is then pushed forward so that the knee is in full contact with the knee brace, with the leg centered in relation to the knee plate. The opposite foot is positioned so it is flush with the instrument and parallel with the foot plate. Next, the knee plate is lowered onto the top of the knee until it is deflected 0.03mm (equivalent to about 2kg of pressure on the knee which yields the smallest degree of error without causing discomfort). The knee height measurement is electronically recorded. Subjects should remain seated between measurements but remove their foot from the foot plate and then return it again. The position of the child in the chair and the foot on the foot plate should be readjusted, and three additional independent measurements should be taken.

The performance of the KHMD has been compared with that of the Valk knemometer on 103 children 6–10y. Growth was measured at 28d intervals with both devices. The within-examiner error for the measurements was 0.295mm for the KHMD and 0.206mm for the knemometer. The correlation between devices for measurements of growth rate on each child was 0.73 ($p < 0.002$) (Cronk et al., 1989).

Knee-height appears to be a valid index of linear growth in well-nourished children. For example, in a study of healthy Indonesian children aged 7–12y, two equations to predict height based on knee height for boys and girls have been developed (Rumapea et al., 2021). However, whether knee height is a valid index of linear growth in children recovering from severe malnutrition is uncertain (Doherty et al., 2001).

10.1.6 Lower leg length in infants and children aged < 3y

Lower leg length (i.e., knee-heel length) is also used to measure short term changes in linear growth over 1–8wks during infancy and early childhood. For preterm infants, normal infants, and toddlers < 3y, a hand-held electronic knemometer (Michaelsen et al., 1991), a mini-knemometer (Hermanussen & Seele, 1997), or an inexpensive vernier or electronic caliper (Skinner et al., 1997; Engström et al., 2003) can be used. Reports for the TEM for these three instruments vary across studies, in part depending on whether the readings are recorded blinded or not, and the number of consecutive readings taken (Hermanussen & Seele, 1997; Skinner et al., 1997; Engström et al., 2003). In general, the inexpensive electronic caliper has a lower TEM than the handheld mini-knemometer (Figure 10.6) and could be used when assessing lower leg length in preterm infants over a short time period (Engström et al., 2003).

The mini-knemometer contains a commercially available electronic slide that discriminates intervals of 10µm. The slide is connected to two measuring arms (A, B) with metallic holders, as shown in (Figure 10.6). The infant's knee and heel are placed between the holders. Knee and heel holders of different sizes can be fitted, depending on the age of the infant.

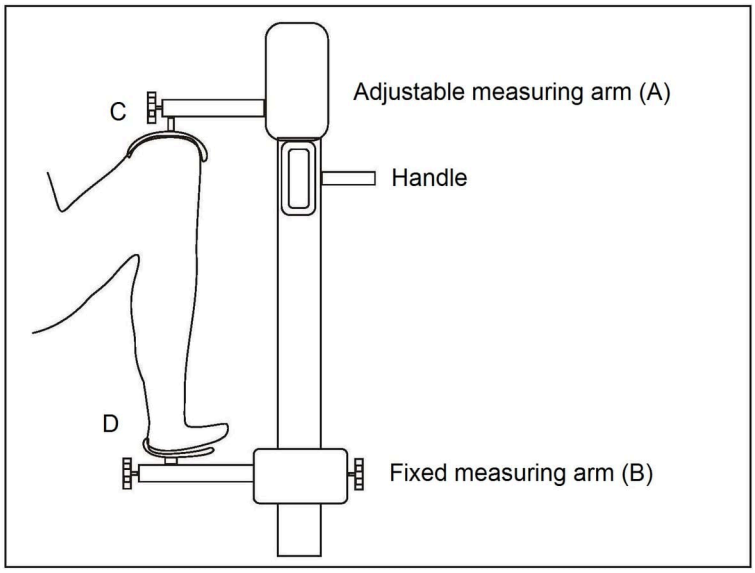


Figure 10.6. Diagram of a mini-knemometer: measuring arms (A and B); metallic holders for knee and heel (C and D). The arms are spring-loaded so that a constant pressure is applied during measurement. Redrawn from Hermanussen and Seele, Annals of Human Biology 24: 307–313, 1997.

When serial measurements are being taken using mini-knemometry, care must be observed to ensure that the infant is always measured in the same position. A spring mounted within the instrument between the arms ensures a constant soft tissue pressure between 2.0 and 3.0N. The measurement is painless for the infant and is best made during breast-feeding. Measurements should be performed without reference to previous recordings. Four–to–five independent measurements of lower leg length should be taken on each child within 1–3min, and both the mean and standard deviation calculated. More readings are needed when the child is restless.

10.1.7 Knee height in adults

Knee height is highly correlated with stature and may be used to estimate height in persons with severe spinal curvature or who are unable to stand. Knee height is measured with a caliper consisting of an adjustable measuring stick with a blade attached to each end at a 90° angle.

Recumbent knee height is measured on the left leg, which is bent at the knee at a 90° angle, while the subject is in a supine

position (Figure 10.7). One of the blades is positioned under the heel of the left foot and the other is placed over the anterior surface of the left thigh above the condyles of the femur and just proximal to the patella. The shaft of the caliper is held parallel to the shaft of the tibia, and gentle pressure is applied to the blades of the caliper.

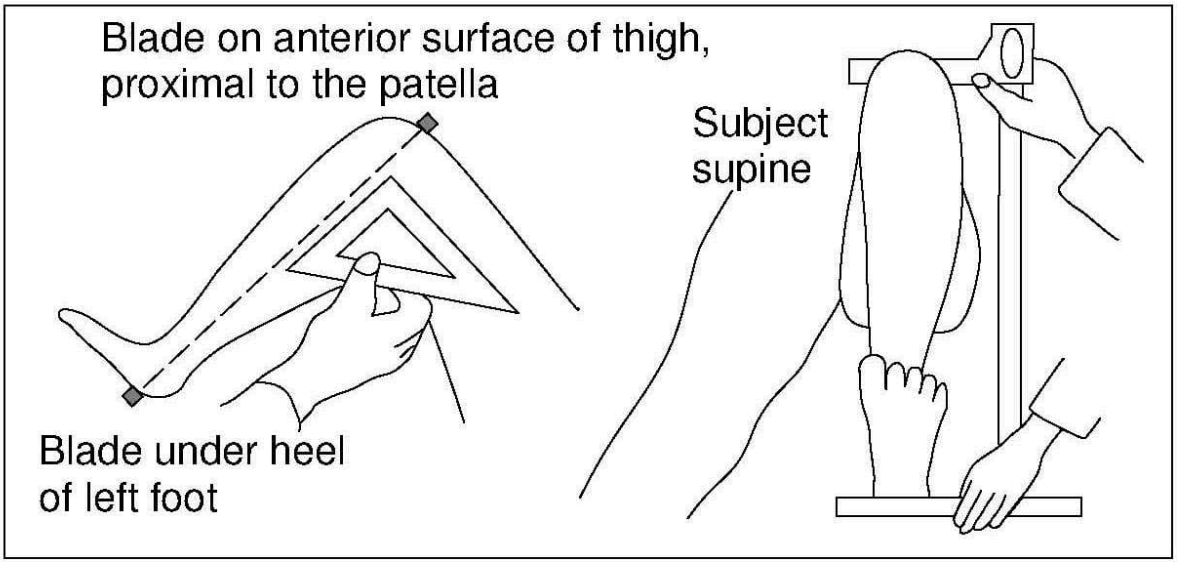


Figure 10.7. Measurement of knee height in adults.

Some of the knee-height calipers are equipped with a locking mechanism to retain the measurement after removing the caliper from the leg. At least two successive measurements should be made, and they should agree within 5mm; the mean is then calculated. Details for measuring knee height in elderly persons seated in wheel chairs are given in (WHO, 1995a).

Formulae are used to estimate adult stature from knee height. Separate stature prediction equations using knee height and age were developed for non-Hispanic white, non-Hispanic black, and Mexican-American elderly persons based on the NHANES III data (Chumlea et al., 1998). However, their appropriateness for estimating stature among other ethnic groups has been questioned. Silva de Lima et al. (2018) compared the validity of 16 equations to estimate height based on knee-height in elderly nursing home residents (n=168) in Brazil. None of the equations examined were applicable for the estimate of height of individuals > 70y, emphasizing that population-specific equations may be necessary. Consequently, an assessment of stature based on knee height should be used only for individuals for whom a direct measurement of stature is not possible, or is likely to be inaccurate, because of vertebral flexions or other skeletal deformities.

10.1.8 Arm span and demi-span

Arm span, like knee height, is also highly correlated with stature and, hence, can be used as an alternative measurement when actual height cannot be used such as in elderly persons when degenerative and osteoporotic changes give rise to spine curvature (Goswami et al., 2018). Arm span is also often used as an alternative to height for the calculation of body mass index (BMI) in older adults (Arlappa et al., 2016), and as a reliable surrogate of both recumbent length and height in healthy children, when these measurements are unobtainable or unreliable (Forman et al., 2014).

The measurement of arm span is easier if carried out against a flat wall (Figure 10.8), to which is attached a fixed marker board at the zero end of a horizontal scale. Sliding on the scale is a vertical movable arm. The horizontal scale should be positioned so that it is just above the shoulders of the subject. Two examiners are needed to measure arm span: one is at the fixed end of the scale; the other positions the movable arm and takes the readings. In elderly persons, an assistant may be needed to support and maintain the arm being assessed in a 90° position (Silva de Lima et al., 2018).

Figure 10.8

Figure 10.8. Measurement of arm span.

For the measurement, the individual should stand with feet together, back against the wall, with the arms extended laterally in

contact with the wall, and with the palms facing forward. The arms should be kept at shoulder height and outstretched maximally. The measurement is taken when the tip of the middle finger (excluding the fingernail) of the right hand is kept in contact with the fixed marker board, while the movable arm is set at the tip of the middle finger (excluding the fingernail) of the left hand. Two readings are taken for each measurement, which is recorded to the nearest 1.0mm (Lohman et al., 1988).

Arm span is difficult to measure in non-ambulatory elderly persons and in individuals with significant chest and spinal deformities and stiffness. (Silva de Lima et al., 2018). Zhang et al. (1998) concluded that in a group of elderly Chinese, for example, knee height provided a more valid estimate of maximum stature during early adulthood than arm span.

Instead of arm span, demi-span is sometimes used to assess body mass index in older adults when an actual height measurement is not possible. Demi-span is the distance between the mid-point of the sternal notch (or jugular notch) — a large visible dip in the neck in humans between the clavicles — and the finger tips, with the right arm outstretched laterally. The measurement is performed using a retractable metal tape and taken to the nearest mm. Again, in some elderly persons, an assistant may be needed to support and maintain the arm being assessed in a 90° position (Silva de Lima et al., 2018).

Demi-span measurements were included in the Health Survey for England (HSE) because they can be measured easily without causing discomfort or distress. New demi-span sex- and age-specific regression equations for estimating adult height have been developed by Hirani and Aresu (2012) and are shown below (standard errors are given in parentheses):

$$\text{Men: } \text{DEH}_{\text{age}}\,(\text{cm}) = 73.0 + 1.30(0.04) \times \text{demi-span} - 0.10 \times \text{age}\ (0.02)$$
$$\text{Women: } \text{DEH}_{\text{age}}\,(\text{cm}) = 85.7 + 1.12(0.05) \times \text{demi-span} - 0.15 \times \text{age}\ (0.05)$$

These equations are based on data for adults ≥ 65y who participated in the HSE. However, as 98% of the sample were white, whether the proposed equations are valid for other ethnic groups is unknown.

10.1.9 Weight in infants and children

In field surveys, a suspended scale and a weighing sling may be used for weighing infants and children < 2y (Figure 10.9). They should be weighed naked or with the minimum of clothing. After slipping the subject into the sling, the weight is recorded as soon as the indicator on the scale has stabilized.

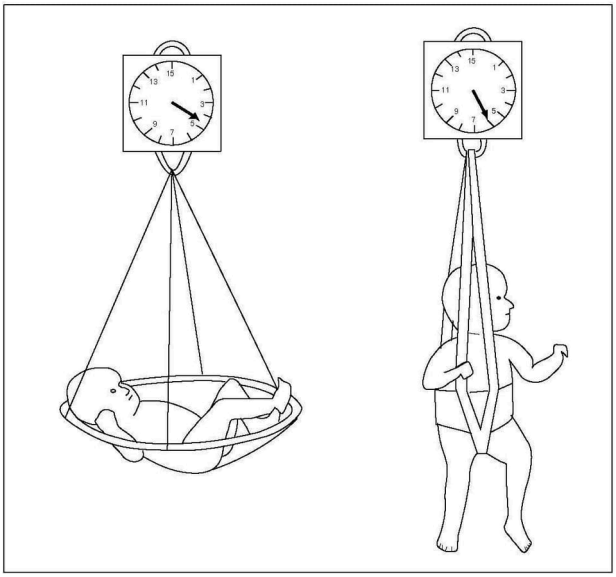


Figure 10.9. Measurement of weight in infants and children using a spring balance. The bowl for the baby is made from a metal or bamboo ring and netting. See comparable devices available from: UNICEF.ORG.

Alternatively, for greater precision, a pediatric scale (within 10g) with a pan may be used (Figure 10.10A). Care must be taken to ensure that the infant (preferably nude) is placed on the pan scale so the weight is distributed equally on each side of the center of the pan. Once the infant is lying quietly, weight is recorded to the nearest 10g. In cold weather, the infant can be wrapped in a

blanket of known weight, which is subtracted afterwards to obtain the weight of the child nude. If the scale has a taring capacity, then the weight of the blanket can be tared before placing the infant on the pan scale.

If there is no alternative, the mother can be weighed alone, and then again holding with the child, using battery-operated precision digital electronic weight scales. The child's weight can then be calculated by subtraction. Alternatively, if the scale has a taring function, then the weight of the mother can be set so the scale readout is zero, enabling only the child's weight to be recorded. If the child cannot be undressed, then standard light clothing of known weight should be worn. This clothing weight should be subtracted afterwards from the child's weight.

10.1.10 Weight in older children and adults

The measurement of weight in older children and adults should be done preferably after the bladder has been emptied and before a meal. A portable digital electronic weight scale, preferably one that has a taring capacity, and has been calibrated to 0.1kg, should be used (Figure 10.10B).

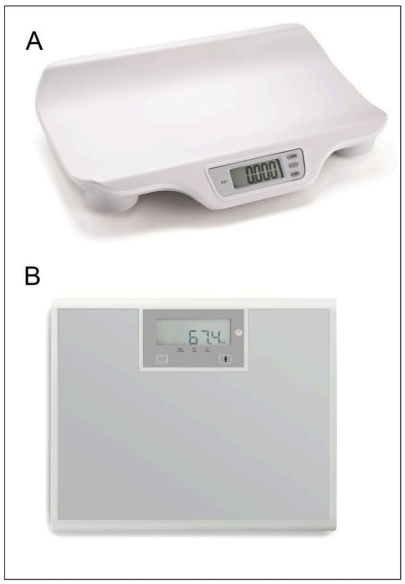


Figure 10.10. Measurement of weight: (A) Digital pediatric scale for infants, (B) Digital balance for a child or adult. Available from:UNICEF.ORG.

The weight scale should be placed on a hard, flat surface (not carpet) and checked and adjusted for zero-balance before each measurement. The subject should stand in the center of the platform and look straight ahead, standing unassisted, relaxed but still, and preferably nude. If nudity is not possible, the subject can wear light underclothing or a paper examination gown, and the weight of these garments should be recorded for later subtraction; standard corrections for clothing should not be used. The presence of visible edema should also be recorded. Body weight should be recorded to the nearest 0.1kg. Again, the time at which the measurement is made should be recorded because diurnal variations in weight occur.

Electronic weight scales should be calibrated with a set of standard weights over the full weight range, both regularly throughout the year and whenever they are moved to another location. Special equipment, such as a movable wheelchair balance beam scale, bed scales, or specialized beds with integrated weighing scales (Mishra et al., 2021) is needed for weighing non-ambulatory persons (Chumlea et al., 1984).

Estimates of weight for the U.S. elderly population can be derived from calf circumference (calf circ), knee height (knee ht), mid-upper-arm circumference (MUAC), and subscapular skinfold (subscap), using equations developed by Chumlea et al. (1989); examples are given below:

$$\text{Weight } (M) = (0.98 \times \text{ calf circ }) + (1.16 \times \text{ knee ht }) + (1.73 \times \text{ MUAC }) + (0.37 \times \text{ subscap }) - 81.69$$
$$\text{Weight } (F) = (1.27 \times \text{ calf circ }) + (0.87 \times \text{ knee ht }) + (0.98 \times \text{ MUAC }) + (0.40 \times \text{ subscap }) - 62.35$$

The above equations were developed from a selected population living in the United States and hence are inappropriate for estimating weights of other populations. Instead, population-specific equations may be required. Quiroz-Olguin et al. (2013) have developed and validated formulae for predicting body weight using circumference-based equations for Mexican adults.

Estimates of weight among multi-racial/ethnic infants and children 0–5.9y in the United States based on ulna length and forearm width and circumference using simple and portable tools have been developed. More investigation of their validity for physically impaired or non-ambulatory children is needed (Zhu et al., 2019).

Use of self-reported weights in adolescents and adults may lead to bias and should be avoided (Ekström et al., 2015; Pérez et al., 2015). In the U.S. Women's Health Initiative, on average women under-reported their weight by about 0.91kg although the discrepancies varied by age, race / ethnicity, education, and body mass index (Luo et al., 2019).

10.1.11 Elbow breadth

Elbow breadth is a good measure of skeletal dimensions and, hence, frame size. The measure is less affected by adiposity than many other anthropometric dimensions and is highly associated with lean body mass and muscle size (Frisancho, 1990).

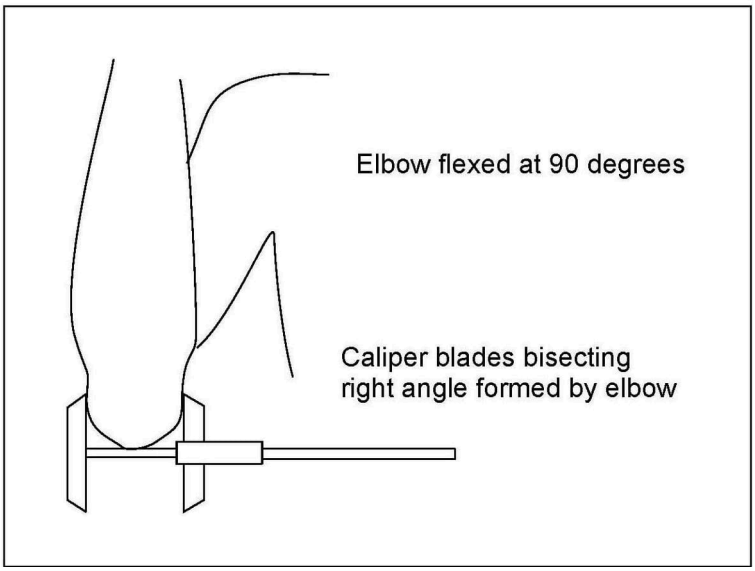


Figure 10.11. Measurement of elbow breadth.

Elbow breadth is measured as the distance between the epicondyles of the humerus. For the measurement, the right arm is raised to the horizontal and the elbow is flexed to 90°, with the back of the hand facing the measurer (Figure 10.11). The measurer then stands in front of the subject and locates the lateral and medial epicondyles of the humerus. The two blades of a flat-bladed sliding caliper are applied to the epicondyles, with the blades pointing upward to bisect the right angle formed at the elbow. Care must be taken to ensure that the caliper is held at a slight angle to the epicondyles and that firm pressure is exerted to minimize the influence of soft tissue on the measurement. The latter is taken to the nearest millimeter (Lohman et al., 1988). Elbow breadth is used to calculate the Frame Index:

This index is used as a measurement of external skeletal robustness in current and past populations (Frisancho, 1990). Studies have confirmed a decline in the Frame Index in recent decades among young children in Germany (Scheffler & Hermanussen, 2014), Russia (Rietsch et al., 2013a), and Argentina (Navazo et al., 2020), a trend that appears to parallel an increase in overweight and obesity. This decline in Frame Index has been associated with a reduction in physical activity, also known to decrease external skeletal robustness (Rietsch et al., 2013b). In 2018, reference percentiles for Frame Index were published for European children and adolescents (Mumm et al., 2018).

This page titled 9.2: Measurements of body size (10.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

9.3: Growth indices, indicators, and recommended growth reference data (10.2)

The correct interpretation and grouping of anthropometric measurements require the use of anthropometric indices (WHO, 1995a). They are usually calculated from two or more raw anthropometric measurements. In the simplest case the indices are numerical ratios such as wt/ht^2 (kg/m^2). Combinations such as weight-for-age, length / height-for-age (i.e., stature-for-age), and weight-for-stature are more complex. These latter growth indices are not ratios and, to avoid confusion with numerical ratios, should not be written as “wt/age”, “ht/age”, and “wt/height”.

Anthropometric indices are often evaluated by comparison with the distribution of appropriate reference data using standard deviation scores (Z-scores) or percentiles. From this, the number and proportion of individuals (as %) with anthropometric indices below or above a predetermined reference limit or cutoff are often calculated. A commonly used statistically defined reference limit for the three main growth indices is a Z-score of –2 (i.e., 2SD below the WHO growth reference median). When used in this way, the index and its associated reference limit become an “indicator”. Growth indicators are often used for public health or socio-medical decision making at the population level. They are also used in clinical settings to identify individuals at risk of malnutrition. Examples of frequently used anthropometric growth indicators and their corresponding application are shown in Table 10.1.

Table 10.1 Anthropometric indicators and their corresponding applications.

Anthropometric indicator	Application
Proportion of children (of defined age and sex) with WHZ < –3	Prevalence of severe wasting
Proportion of children (of defined age and sex) with WHZ < –2	Prevalence of wasting
Proportion of children (of defined age and sex) with WHZ > +2	Prevalence of overweight
Proportion of children (of defined age and sex) with HAZ < –2	Prevalence of stunting
Proportion of children (of defined age and sex) with WAZ < –2	Prevalence of underweight
Proportion of children (of defined age and sex) with BMIZ +1 to +2	Prevalence of “at risk of over-weight”(for those 0–5y)
Proportion of children (of defined age and sex) with BMIZ > +2	Prevalence of overweight (for those aged 0–5y)
Proportion of children (of defined age and sex) with BMIZ > +3	Prevalence of obesity (for those aged 0–5y)

The WHO recommends the use of the WHO Child Growth Standards for young children from birth to 5y for international use (Growth Standards), given the small effect of ethnic and genetic differences on the growth of infants and young children compared with the environmental, nutritional, and socio-economic effects, some of which may persist across generations. These new growth standards were developed because of the technical and biological limitations identified with the earlier NCHS/WHO growth reference; for more details, see Garza & de Onis, (1999).

A prescriptive approach depicting normal early childhood growth under optimal environmental conditions was used for the new WHO Child Growth Standards so they represent how young children should grow, rather than as a “reference” describing how children do grow. For details of the methods used in the compilation of the new WHO Child Growth Standards, see de Onis et al. (2004).

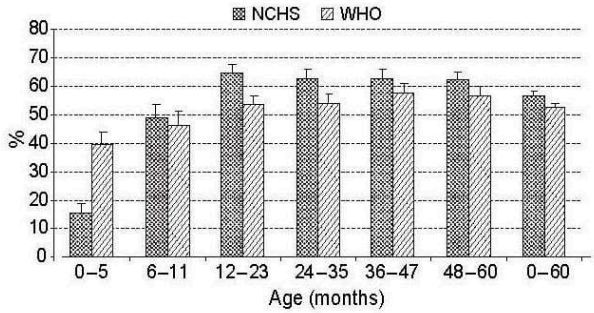
WHO has developed a tool for the application of the WHO Child Growth Standards which includes instructions on how to take the measurements, interpret growth indicators, investigate causes of growth problems, and how to counsel caregivers. For more details see: WHO Child Growth Training Module (Training Module).

For boys and girls charts (0–6mos; 0–2y; 6mos–2y; 0–5y) for head circumference-for-age, length/ height-for-age, weight-for-age, weight-for-length (45–110cm), weight-for-height (65–120cm), and body mass index (BMI)-for-age are available, expressed as percentiles or Z-scores. Data tables for Z-scores and percentiles can also be downloaded from WHO (Growth Standards).

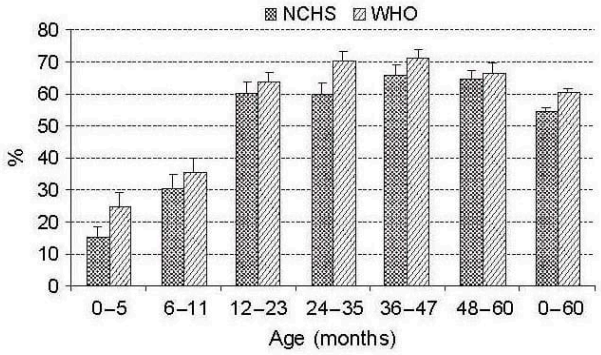
The WHO growth reference for school-age children and adolescents 5–19y (or 61–228mos) is recommended for international use (de Onis et al., 2007). This is based on a reconstruction of the original U.S. National Centre for Health Statistics (NCHS) data set supplemented with data for under-fives from the WHO Child Growth Standard. The statistical methodology used to construct this reference was the same as that used for the WHO Child Growth Standard. For boys and girls charts for height-for-age (5–19y), BMI-for-age (5–19y) and weight-for-age (5–10y) are available, expressed as Z-scores or percentiles. Data tables for Z-scores and percentiles are also available (Growth Reference Data).

Concern over the impact of using the growth charts based on the new WHO Child Growth Standard compared to the earlier NCHS/ WHO international growth charts on the estimates of malnutrition for children from birth to age 5y prompted investigations by several researchers (de Onis et al., 2006;Martorell & Young, 2012). Such information is important because the resulting pattern of child growth, following an intervention designed to improve child health, depends on the growth charts used. Based on the growth indices outlined above, de Onis and co-workers (2006) compared the rates of underweight, stunting, wasting, severe wasting, and overweight generated from the two sets of growth charts using two cross-sectional datasets for children 0–5y representative of different nutritional status profiles from Bangladesh and the Dominican Republic (for overweight only). Important differences were revealed between the WHO standard and the NCHS reference that varied by age group, growth indicator, and the nutritional status of the populations examined. These same differences in the age-related patterns of stunting and wasting have also been reported between the two growth reference data when compared using nationally representative data from India and Guatemala (Martorell & Young, 2012).

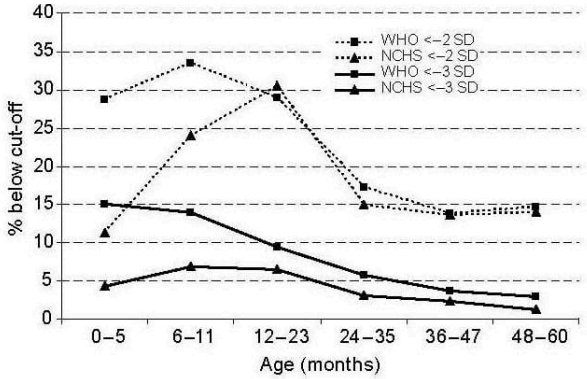
In general, for populations with nutritional status profiles similar to those of Bangladesh, rates of underweight (Figure 10.12) during the first half of infancy, and stunting (Figure 10.13) at all ages, will be higher when based on the WHO standard compared with those of the NCHS reference, whereas after the first six months, rates for underweight will be lower. For wasting (i.e., weight-for-stature Z-core < –2), the rates will be greater at all ages, with the difference being especially striking during early infancy (0–6mos), as shown in Figure 10.14. A similar pattern exists for severe wasting (i.e., weight-for-length/ height Z-score < –3) and applied as the criterion for enrolling children in therapeutic feeding programs (WHO/UNICEF, 2009).



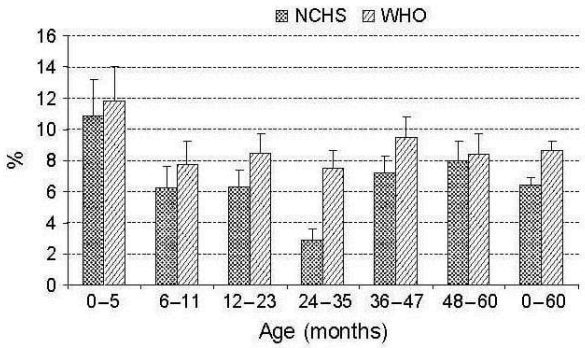
10.12 Prevalence of underweight (below −2 standard deviations from the median for weight-for-age) by age based on the World Health Organization (WHO) standards and the National Center for Health Statistics (NCHS) reference in Bangladesh



10.13 Prevalence of stunting (below −2 standard deviations from the median for length/height-for-age) by age based on the World Health Organization (WHO) standards and the National Center for Health Statistics (NCHS) reference in Bangladesh



10.14 Prevalence of wasting (below −2 standard deviations (SD) from the median for weight-for-length/height) and severe wasting (below −3SD from the median for weight-for-length/height) by age based on the World Health Organization (WHO) standards and the National Center for Health Statistics (NCHS) reference in Bangladesh



10.15 Prevalence of overweight (above +2 standard deviations from the median for weight-for-length/height) by age based on the World Health Organization (WHO) standards and the National Center for Health Statistics (NCHS) reference in the Dominican Republic

Figs 10.12 – 10.15. Prevalence of underweight, stunting, wasting, and overweight, showing the differences between the WHO and NCHS standards. From: de Onis et al. (2006).

For overweight (i.e., weight-for-stature > +2 Z-scores), the expected difference in overweight for children 0–60mos assessed by the WHO standard will be an increase in the rates of overweight as shown in Figure 10.15, although the pattern varies by age group, and will also vary with the nutritional status of the population under study. Differences in overweight defined by BMI-for-age Z-scores could not be evaluated because BMI-for-age data are not available for preschool children in the older NCHS growth reference.

The reported differences in the age patterns of growth between the two growth charts, particularly during infancy, are likely because the WHO standard represents the growth of only term breast-fed infants who have greater weights for length than the predominantly formula-fed Caucasian infants of the NCHS reference, although they do become thinner during the second year and beyond (de Onis et al., 2006). An additional factor is the shorter intervals between measurements during the periods of rapid growth for the WHO standards compared to the NCHS reference (de Onis et al., 2006).

The U.S. have also released updated CDC 2000 childhood (growth charts), based primarily on physical measurements taken during five nationally representative surveys conducted between 1963 and 1994, although some supplemental data were also used. During the creation of these revised CDC 2000 Growth Charts, two data sets were excluded: growth data for very low birth-weight (VLBW) infants (< 1500g) whose growth differs from that of normal birthweight infants, and weight data for children > 6y

who participated in the NHANES III survey. The latter data were excluded from both the revised weight and BMI growth charts because the inclusion of these data shifted the upper percentile curves. With the exclusion of these selected data, a modified growth reference that is not a purely descriptive growth reference was created because it does not contain representative national data for all variables (Kuczmarski et al., 2002). Users can compute exact percentiles, Z-scores, and BMI values (from weight and data on length or height) using WHO Anthro software and the CDC reference. For more details, see Kuczmarski et al. (2002). A comparison of these CDC 2000 Growth Charts with the WHO Child Growth Standards is available in de Onis et al. (2006).

A series of prescriptive standards are now available for monitoring fetal, newborn growth, and gestational weight gain for international use developed by the INTERGROWTH-21st Project. As noted earlier, this project adhered to the WHO recommendations for assessing human size and growth and followed healthy pregnant women longitudinally from 9wks of fetal life to 2y (Papageorghiou et al., 2018). Box 10.1. lists the prescriptive standards generated from the INTERGROWTH-21st project with the associated references.

st international standards for monitoring growth and development from early pregnancy to 2y.

- Fetal growth standards based on serial ultrasound measurements: Papageorghio et al. (2014).
- Estimated fetal weight standards: Stirnemann et al. (2017).
- Fetal growth velocity standards from the Fetal Growth longitudinal Study: Ohuma et al. (2021).
- Gestational weight gain standards based on women enrolled in Fetal Growth longitudinal study: Ismail et al. (2016).
- Newborn weight, length, and head circumference by gestational age and sex: Villar et al. (2014).
- Postnatal growth standards for preterm infants: Villar et al. (2015).

Of the indicators listed in Table 10.1, three are also included as components of the six Global Nutrition Targets for 2030 set by WHO/UNICEF and shown in Box 10.2.

Box 10.2 Global nutrition targets for 2030

- Stunting. Target: 50% reduction in the number of children under 5 who are stunted.
- Anemia. Target: 50% reduction of anemia in women of reproductive age.
- Low birthweight. Target: 30% reduction in low birthweight.
- Childhood overweight. Target: Reduce and maintain childhood overweight to less than 3%.
- Breastfeeding. Target: Increase the rate of breastfeeding in the first 6 months up to at least 70%.
- Wasting. Target: Reduce and maintain childhood wasting to less than 3%.

From WHO/UNICEF (2021),

Policy briefs provide details of the extensions of each of the 2025 Global Nutrition Targets. WHO has also developed a web-based tracking tool to assist countries to set national targets and chart progress for achieving the six global targets (See Section 10.2.6).

Additional factors that must be considered when selecting an index or combination of indices to evaluate growth include the availability of accurate measuring equipment, the training of examiners to collect accurate information and to interpret the results correctly, and the time required to take the measurements. Finally, often overlooked are the costs of not identifying undernourished children or incorrectly identifying adequately nourished children as undernourished (Gorstein et al., 1994).

Details of the growth indices, the indicators derived from them and their associated applications, together with their advantages and limitations, are discussed below. Information on the available interpretive criteria for each growth index is also included. .

10.2.1 Head circumference-for-age

Intrauterine growth retardation, or chronic malnutrition during the first few months of life, may hinder brain development and result in an abnormally low head circumference. Hence, head circumference is a widely used proxy of neural growth and brain size. When brain size is outside of normal values, it is an important risk factor for cognitive and motor delay.

Head circumference-for-age can be used as an index of chronic malnutrition for children < 2y but is not sensitive to less extreme

malnutrition(Yarbrough et al., 1974). Beyond age 2y, growth in head circumference is slow and its measurement is no longer useful, so an indicator based on head circumference-for-age is not included in Table 10.1 (Harris, 2015). Certain non-nutritional factors, including some diseases and pathological conditions (e.g., microcephaly), genetic variation, and cultural practices such as binding of the head during infancy, as well as a difficult or forceps-assisted delivery at birth, may also influence head circumference.

Interpretive criteria

Microcephaly in an infant is defined as a measurement of head circumference that is more than 2SD below the mean of an age- and sex-appropriate growth chart, whereas in severe microcephaly, head circumference is more than 3SD below the mean (Harris, 2015). For international use, the WHO Child Growth Standards that include head circumference by age and sex from birth to 13wks, birth to 2y, and birth to 5y are recommended (Growth Standards). These standards, unlike the INTERGROWTH-21st standards described below, only include term newborn infants.

Papageorghiou et al. (2014) have produced international standards for fetal growth in which head circumference was measured using ultrasound from 14wks to 42wks gestation. The 3rd, 5th, 10th, 50th, 90th, 95th, and 97th smoothed percentile curves for fetal head circumference are available.

International standards for newborn head circumference by gestational age (33–42wks) and sex are also available from the INTERGROWTH-21stProject. Pregnancies of all women who met strict eligibility criteria for a population at low risk of impaired fetal growth were selected and followed prospectively. Gestational age was estimated using ultrasound. Hence, these multi-ethnic growth standards represent how fetuses should grow (i.e., the standards are prescriptive) and can be used to diagnose fetal growth restrictions world-wide and allow comparisons of newborn size across multi-ethnic populations (Villar et al., 2014). Free software is available through the INTERGROWTH-21st Project website to calculate Z-scores and centiles.

10.2.2 Weight-for-age

Weight-for-age reflects body mass relative to chronological age. Low weight-for-age is described as “lightness” and reflects a pathological process referred to as “underweight”, arising from gaining insufficient weight relative to age, or losing weight (WHO, 1995a). Because of its simplicity and the availability of scales in most health centers in low- and middle-income countries, weight-for-age is widely used in children from 6mo to 7y to assess underweight.

Underweight is defined by the indicator weight-for-age < –2 Z-score (relative to the WHO Child Growth Standard) and was selected as one of the indicators to track the progress in addressing poverty and hunger for the UN Millennium Development Goals (MDGs). These are now replaced by the UN Sustainable Development Goals (SDGs) (2015–2030), which include the indicators to monitor progress shown in Box 10.2.

A major limitation of weight-for-age is that it is influenced by both the height and weight of a child, making interpretation difficult. For example, weight-for-age fails to distinguish tall thin children (i.e., those with low weight-for-height) from those who are short (i.e., low height-for-age) but with adequate weight. As a result, the use of weight-for-age alone to estimate the prevalence of undernutrition leads to a gross underestimate of the problem in populations where the prevalence of low height-for-age (i.e., linear growth retardation) is high but that of low weight-for-height (i.e., wasting) is low (e.g., Guatemala) (WHO, 1995a). Conversely, in countries undergoing the nutrition transition and experiencing progressive increases in childhood overweight and obesity, use of weight-for-age alone will result in overstating their progress in reducing underweight and mask stunting (Uauy et al., 2008).

To interpret any single measurement of weight (or height) in relation to the reference data, the exact age of the child at the date of the measurement must be calculated from the date of birth. Software, such as WHO AnthroPlus, can calculate exact ages in decimal fractions of a year, from birth and visit dates (AnthroPlus 2009). Details of WHO Anthro Growth Standards are given here (Growth Standards).

Even when information on the date of birth is available, ages are sometimes reported following rounding off the most recently attained whole month. This practice should not be followed, because it results in systematic errors (Gorstein et al., 1989). In the event that documentary evidence of the date of birth is not available, it may be necessary to obtain at least the month and year of birth using a local events calendar. Information on the development of a local events calendar is available (FAO, 2008). Details of other methods that can be used to assess the age of children are given in Chapter 9.

Birthweight is also measured in health centers in low-income countries as well as maternity hospitals because it is an important

indicator of fetal and neonatal health. Low birthweight (LBW) is defined as < 2500g at birth, and very low birthweight as < 1500g. South Asia is a region where 25% of births are LBW, which is of concern given that LBW infants are at risk of poor health and development outcomes. Consequently, as noted earlier, one of the Global Nutrition Targets is to achieve a 30% reduction in low birthweight by 2030 in recognition of the importance of LBW for survival, development and health in the lifespan (WHO, 2014).

Low birthweight may be a consequence of premature birth (i.e., before 37wks gestation), being intrauterine growth restricted (IUGR), or both. To assess fetal growth restriction, the weight of the newborn is compared to the achievement of the expected weight for a given gestational age. Small for gestational age (SGA), defined as being born below the 10th percentile of a sex-specific birthweight distribution at a specified gestational age, is often used as a proxy to identify IUGR neonates. Methods for estimating gestational age have been described in Section 10.1.2.

Interpretive criteria

For international use, data for weight-for-age, expressed as Z-scores and percentiles, are available for term infants and children (0–6mos; 0–2y; 6mo–2y; 2–5y; 0–5y) based on the WHO Child (Growth Standards), as noted earlier. These data only include term newborn infants (i.e., not gestational age specific at birth), unlike the birthweight data of the INTERGROWTH 21st Project (Villar et al. 2014).

Data for weight-for-age for children beyond age 10y are not available (de Onis et al., 2007) because beyond age 10y, children are experiencing the pubertal growth spurt and may appear to have excess weight based on weight-for-age when in fact they are just tall. The U.S. CDC 2000 growth reference data, however, provide weight-for-age growth charts for boys and girls from birth to 36mos and 2–20y (Kuczmarski et al., 2002). As emphasized earlier (Section 10.2), weight data for children > 6y who participated in NHANES III survey were excluded from these revised CDC 2000 weight growth charts because the inclusion of these data shifted the upper percentile curves.

Differences in the prevalence of underweight (i.e., weight-for-age Z-score < –2) according to the WHO standard and the older NCHS reference have been reported based on the same dataset, as noted in Section 10.2. Such differences have affected the ranking of countries with respect to underweight. For example, using the 1996–1997 DHS survey from Bangladesh, the prevalence of underweight was much higher (i.e., 2.5 times greater) during the first six months but lower thereafter when based on the WHO standard compared to the NCHS reference (de Onis et al., 2006). As noted earlier (Section 10.2), this difference has arisen because the WHO standard is based on breast-fed infants whose growth pattern in infancy differs substantially from that of the predominantly formula-fed infants of the NCHS reference (de Onis et al., 2006).

International standards for optimal fetal growth and newborn weight (and length and head circumference) by sex and gestational-age (33–42wks) are available from the INTERGROWTH-21st Project (Villar et al., 2014; Papageorghiou et al., 2014). Reliable estimates of gestational age were obtained by ultrasound measurements and newborn anthropometric measurements obtained within 12h of birth. These international fetal growth standards should be used worldwide to diagnose fetal growth restriction uniformly and monitor growth from early pregnancy through to the neonatal period(Papageorghiou et al., 2018), rather than locally-produced reference data. Use of these international fecal growth standards derived from a healthy population reduces the risk of under-diagnosing fetal growth restriction, which may occur when locally-produced reference data that includes high-risk mothers are used. The international standard for weight by gestational age and sex for newborn infants allows the accurate assessment of the prevalence of SGA infants worldwide. The 3rd, 10th, 50th, 90th, and 97th smoothed percentile curves and the numerical values for birthweight according to gestational age (33 to 43wks) are available (Villar et al., 2014). For more details of the INTERGROWTH-21st Project, see Box 10.1.

10.2.3 Weight-for-stature

Weight-for-stature, often referred to as “Weight-for-length/height”, measures body weight relative to length or height. Low weight-for-stature in children is described as “thinness” and reflects a pathological process referred to as “wasting”. It arises from a failure to gain sufficient weight relative to stature or from losing weight. High weight-for-stature in children is termed “overweight” and arises from gaining excess weight relative to length or height or from gaining insufficient length or height relative to weight (WHO, 1995a).

The prevalence of wasting is defined as the proportion of children with a Z-score for weight-for-stature < –2 (i.e., below –2SD from the WHO median for weight-for-length/height). Wasting often develops very rapidly, and is associated with both changes in

the food supply and the prevalence of infectious diseases. Wasting can be reversed quickly with an appropriate intervention. As a result, weight-for-stature is the preferred anthropometric index for identifying young children who are most likely to benefit from a feeding program, or for evaluating the benefits of intervention programs. It is more sensitive to changes in nutritional status than stature-for-age.

A Z-score between –2 and –3 is often defined as "moderate acute malnutrition". Children with moderate acute malnutrition have specific nutrient needs, details of which are available in Golden (2009). When the Z-score for weight-for-stature falls below –3, severe acute malnutrition (SAM) is present (WHO / UNICEF, 2009). Children are sometimes discharged following treatment for SAM when weight-for-stature reaches a Z-score of > –2 relative to the WHO Child (Growth Standards) with no pitting edema (WHO, 2013). This discharge criterion is based on the lower risk of mortality reported compared to those children with a Z-score more negative than –2.0.

In the past, the peak prevalence of wasting was said to occur in children at aged 12–23mos, triggered by inappropriate complementary feeding and the depleting effects of infectious diseases, particularly diarrhea (WHO, 1986). However, these early studies were based on the use of the NCHS reference in which the majority of the infants were formula-fed, with a different growth pattern compared to their breast-fed counterparts, as noted earlier. As a result, use of the new WHO Child Growth Standard based on breast-fed infants has revealed an alarming level of wasting during infancy, particularly in South Asian countries, a trend that was masked when the earlier NCHS reference was applied, as shown in Figure 10.16.

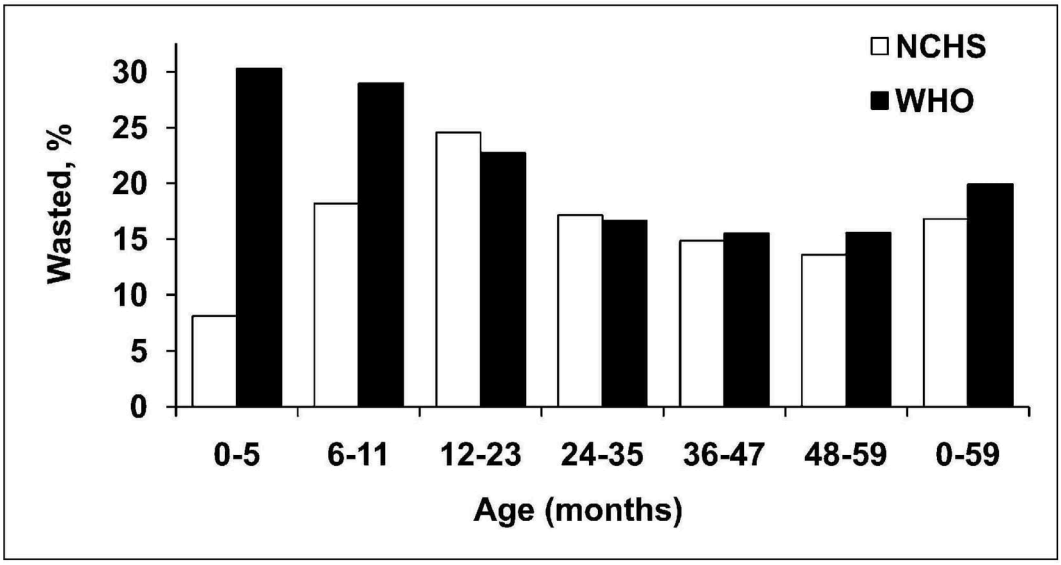


Figure 10.16. Percentage of children (N=41,306) with wasting (weight-for-length / height Z-score < –2) by age and reference (NCHS and WHO) in India. Data are from the Indian National Family Health Survey 3 (2005–2006). Redrawn from Martorell and Young (2012)

As discussed earlier in Section 10.2, the striking differences in the prevalence of wasting are likely due to the inclusion of only breast-fed infants in the WHO sample, whereas the NCHS reference was based predominately on formula-fed children, as noted earlier (see Figure 10.14). Differences in the measurement intervals used may also be a factor.

Numerous preconceptual / prenatal and postnatal factors have been associated with wasting. Of the preconceptual / prenatal factors, short maternal stature, low BMI, and intrauterine growth restriction are well recognized (Black et al., 2008), whereas maternal depression has only recently been explored (Ashaba et al., 2015). Postnatal attributable factors comprise poor infant and young child feeding, recurrent infections, and small body size in early life. In some circumstances, household wealth and sub-optimal child care practices have also been identified (Martorell & Young, 2012).

Seasonality is another factor influencing the prevalence of wasting. Rates are often higher in the rainy season when staple foods

stored from the previous year's harvest may be depleted, morbidity is increased, and there is greater participation by women in the labor market (Hillbruner & Egan, 2008; Baye & Hirvonen, 2020).

Because the fifth Global Nutrition Target for 2030 (Box 10.2) includes “reduce and maintain childhood wasting to less than 3%” (WHO/UNICEF, 2021), WHO and UNICEF have developed a screening tool that includes five prevalence thresholds for wasting for global monitoring and for identifying priority countries for action; see Section 10.2.6 for more discussion of this tool.

Studies using *in vivo* laboratory methods of body composition have shown that wasting is associated with major deficits in both the fat-free mass and fat mass. The methods employed include deuterium dilution (D_2O), dual-energy X-ray absorptiometry (DXA), or bioelectrical impedance analysis (BIA) (Chapter 14). In severe wasting, the extent of the decline in fat-free mass appears to be in proportion to the severity of wasting, whereas the decline in fat mass is more modest. Moreover, following treatment, although levels of fat may recover, those of fat-free mass remain low in the longer-term even following treatment; this pattern is shown in Figure 10.17

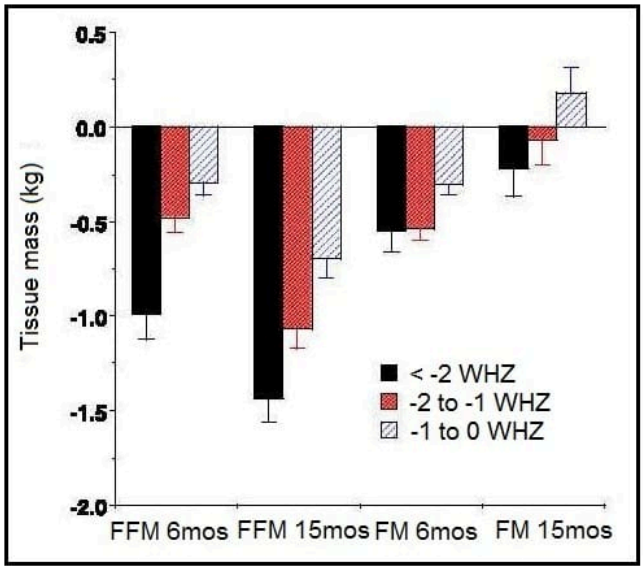


Figure 10.17. Association between body composition outcomes expressed as absolute fat-free mass (FFM) and fat mass (FM) and the severity of wasting, categorized by weight-for-height scores. Effects for the three categories of low weight-for-height z-score (WHZ) are expressed relative to a reference group with WHZ > 0. Redrawn from Wells (2019)

with data from a randomized trial in Cambodia in which infants received four ready-to-use therapeutic foods (RUTFS) from age 6–15mos. Body composition in this study was assessed using deuterium dilution (D_2O) at 6mos and 15mos of age (Skau et al., 2019). For a detailed review of body composition in undernourished children, see Wells (2019)

Overweight and obesity in young children is becoming increasingly common worldwide(de Onis & Lobstein, 2010). Recognition of the role of childhood overweight and obesity, and their subsequent causal link with diabetes and other chronic diseases in adulthood, has led WHO to include prevalence thresholds for overweight in children < 5y (defined as weight-for-height > 2 Z-score) as well as wasting. Moreover, WHO/ UNICEF has endorsed “reduce and maintain childhood overweight to less than 3% in 2030” as one of the six Global Nutrition Targets shown in Box 10.2 (WHO/UNICEF, 2021).

Differences also arise in the prevalence of overweight (defined as > +2 Z-scores weight-for-stature for children), with a higher prevalence for all age groups when estimated by the WHO standard compared to the older NCHS reference, as noted in Section 10.2; for more discussion see (de Onis et al., 2006).

Interpretive standards

WHO Child Growth Standards provide data for weight-for-length for boys and girls age 0–2y; 2–5y; 0–5y expressed as Z-scores or percentiles and available as charts or tables (Growth Standards). However, during adolescence, the weight-for-height relationship

changes dramatically with age and also with maturational status. For this reason, the WHO growth reference for school-age children and adolescents 5–19y does not include weight-for-height reference data. Instead, they include BMI-for-age (5–19y) Z-scores and percentiles (de Onis et al., 2007). The full set of charts and tables displayed by sex and age (years and months), percentile and Z-score values and related information are available (Growth Reference Data).

The U.S. CDC 2000 growth charts are available for weight-for-length data by sex from birth to 36mos, as well as weight-for-height reference data for prepubescent boys and girls that extend from a height of 77–121cm (growth charts),

Note that increasingly, body mass index-for-age (BMI), expressed as a Z-score or as a percentile, is being used to assess overweight and obesity in both childhood (0–5y) and in school children and adolescents (5–19y). The BMI indicator is replacing the index weight-for-stature used in the past to define overweight and obesity in childhood (0–5y). The WHO defines a child 0–5y as "at risk of overweight" if BMI Z-score is > +1 and < +2; "overweight" if BMI Z-score > +2 and "obese" if BMI Z-score is > +3, based on the WHO Child Growth Standard (de Onis & Lobstein, 2010). For children aged 5–19y, however, BMI-for-age Z-scores above +1 and +2 based on the WHO 2007 growth reference data (de Onis et al., 2007) are applied to define overweight and obesity, respectively. In contrast, based on the U.S. CDC 2000 growth charts, overweight in a child is defined as a BMI > 85thpercentile but < 95th percentile for age and gender, whereas a BMI > 95th percentile is indicative of obesity;

The CDC2000 BMI-for-age growth charts did not extend beyond the 97th percentile. With the increasing prevalence of severe obesity in the United States, CDC has developed new percentiles to monitor very high BMI values by including data on children and adolescents from 1988–2016. These extended BMI percentiles include the 95th, 98th, 99th, and 99.9th percentiles by age (2–20y) and sex (Extended BMI-for-Age Charts).

10.2.4 Length or height-for-age

Length or height-for-age (i.e., stature-for-age) is a measure of achieved linear growth that can be used as an index of past nutritional or health status. Recumbent length is measured in infants and children less than 2y, and height in older children. Low height-for-age is defined as "shortness" and reflects either a normal variation or a pathological process involving failure to reach linear growth potential. The outcome of the latter process is the gaining of insufficient height relative to age and is referred to as linear growth retardation (or linear growth faltering) (Leroy & Frongillo, 2019).

The number of children suffering from linear growth retardation is much higher than the number of children who are stunted (Roth et al., 2017). Stunting is defined as having a length or height-for-age Z-score < –2SD based on the WHO Child Growth Standard for children age 0–5y (Growth Standards). When the prevalence of stunting is greater than 40%, it is considered a severe public health problem (WHO, 1995a). In populations with a high prevalence of stunting, usually the entire height distribution has shifted downward, suggesting that most, if not all individuals, have been affected and are not growing to their full potential. However, when the prevalence is much lower and approximates the expected level (i.e., ~ 2.5% of a healthy population), then those with low height-for-age are likely to be genetically short. WHO/UNICEF, 2021 has set a 50% reduction in stunting for children <5y by 2030 as one of the global nutrition targets (Box 10.2).

The prevalence of linear growth retardation generally peaks during the second or third year of life (Roth et al., 2017; Leroy et al., 2015). This age pattern is much less affected by the growth reference data used than wasting (de Onis et al., 2006), although the prevalence estimates for stunting do appear to be higher based on the WHO Child Growth Standard compared to the older NCHS reference data. The difference is attributed to the tighter variability of the WHO standards which affects the placement of the usual reference limit for stunting (i.e., –2SD)(de Onis et al., 2006). These discrepancies in the trends for stunting are evident in Figure 10.18, based on data from India. Here, the greatest difference between the WHO and NCHS reference is apparent in the 0–5mos age group (i.e., 20% vs. 10%).

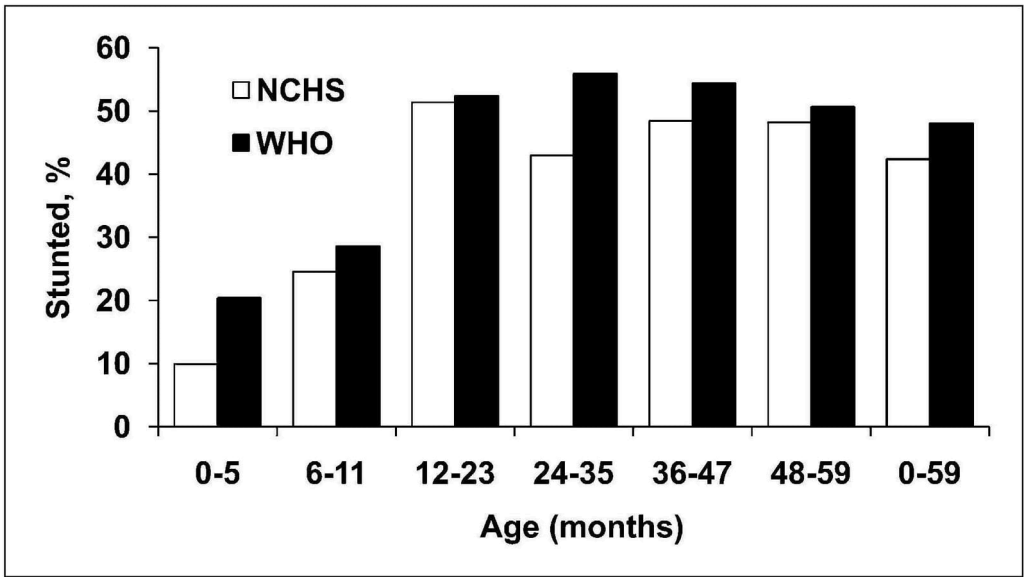


Figure 10.18. Percentage of children (N=41,306) with stunting (length-for-age < –2Z-score by age and reference (NCHS and WHO) in India (National Family Health Survey, 2005–2006). Redrawn from Martorell and Young (2012).

Unfortunately, in many low-income countries, length at birth or in the neonatal period, is seldom reported because of the difficulty of recording at the time of birth and the lack of the appropriate equipment and measurement training (Solomons et al., 2015). As a result, assessment of rates of linear growth retardation in many low-income countries are often based on those for children aged from 6–59mos, so the age of onset of growth failure is uncertain.

In a recent longitudinal birth cohort study, length (n=1197) was measured shortly after birth (mean 7.7d post-delivery) in 7 resource-poor settings in Bangladesh, Brazil, India, Nepal, Peru, South Africa, and Tanzania; gestational age was not assessed (MAL-ED Network Investigators, 2017). Here, shortly after birth the prevalence of a length-for-age Z-score below –1 of the WHO Child Growth Standard was 43% (range 37%–47% across sites), and that of stunting (i.e., length-for-age Z-score below –2) 13% (range 10%–16% across sites).

It is of interest that in the MAL-ED longitudinal birth cohort, there was an almost uniform decrease in length-for-age with age (except for Brazil), with the greatest increase in stunting in all sites (except Brazil) occurring after 6mos. The prevalence of stunting reached a plateau at about two years. Hence, despite diverse cultures and geography, the patterns of stunting with age were the same across study sites (except for Brazil), although the magnitude differed, as shown in Figure 10.19.

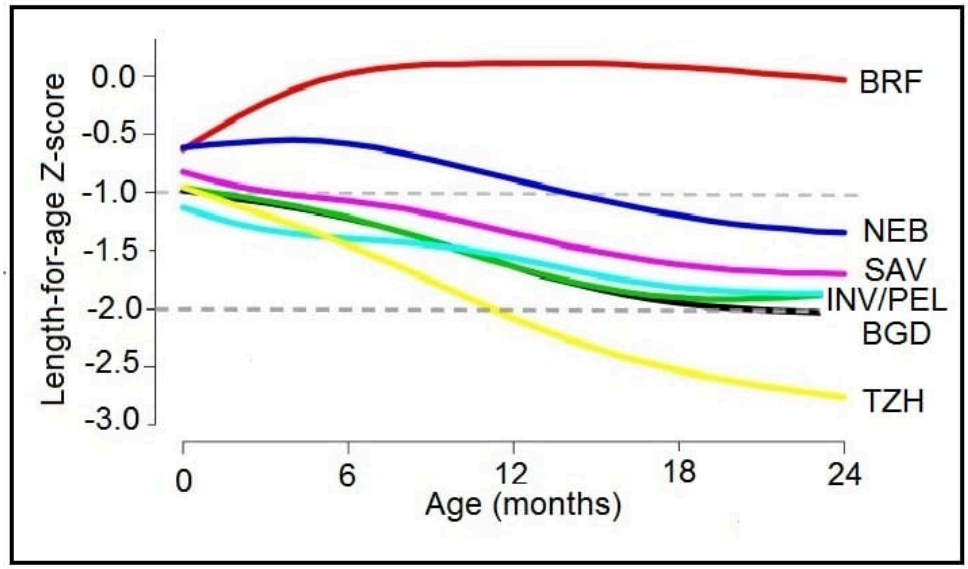


Figure 10.19. Length-for-age stratified by age and site. Site-specific length-for-age trajectories with age were smoothed using a smoothing spline. Sites include BGD, Dhaka, Bangladesh; BRF, Fortaleza, Brazil; INV, Vellore, India; NEB, Bhaktapur, Nepal; PEL, Loreto, Peru; SAV, Venda, South Africa; TZH, Haydom, Tanzania. Redrawn from MAL-ED study (2017)

The stunting rates at birth reported in the MAL-ED cohort study are in striking contrast to the stunting prevalence within 6wks of birth reported for young infants of Mayan indigenous origin in Guatemala. Here stunting rates were 36-47% indicating that impaired fetal growth was the major predictor of early infant linear growth failure among these infants of Mayan indigenous origin (Bernagard et al., 2013; Solomons et al., 2015).

The etiology of linear growth faltering is multifactorial, and is associated with many of the same prenatal and postnatal factors identified for wasting (Section 10.2.3). For example, in the MAL-ED longitudinal birth cohort, five factors were determinants of linear growth faltering during early childhood: lower enrolment weight, shorter maternal height, higher prevalence of entero-pathogen detection, lower socioeconomic status, and consumption of a lower percent of energy from protein in non-breast-milk foods(MAL-ED Network Investigators, 2017). For a comprehensive review of the factors recognized as causal for stunting, see Bhutta et al. (2008).

Interestingly, some of the predictors of stunting may not be the same as those linked to the recovery from this condition. In the MAL-ED longitudinal cohort, the timing of stunting was significantly associated with recovery from stunting. Their findings suggested children who are stunted at an older age have a higher chance of recovering from stunted growth than those children who are stunted earlier, i.e, at age 6mos. (Das et al., 2021).

Identifying and preventing linear growth retardation or stunting during childhood is important because impaired linear growth results in a reduction in adult size, which, in turn, is causally linked with difficulties in child birth and poor birth outcomes, notably increased risk of low-birthweight infants(Sinha et al., 2018). Delays in cognitive and motor development in childhood, reduced earnings in adulthood, and chronic diseases have also been presented as consequences of linear growth retardation and stunting during early childhood. Based on the current evidence, however, these outcomes are unlikely to be causally related, but instead correlates of linear growth retardation and stunting (Leroy & Frongillo, 2019).

Several early cross-sectional studies have linked stunting with an elevated risk of overweight in children when classified by BMI or weight-for-height(Popkin et al., 1996; Sawaya et al., 1995), although longitudinal studies have reported contrasting findings (Walker et al., 2007; Kagura et al., 2012). Caution is needed when using BMI (and weight-for-height) to examine associations between stunting and body composition, given that height is incorporated in the measurement of both stunting (HAZ) and BMI. Wells (2019) emphasizes that this will generate an autocorrelation between short stature and high BMI. Instead, it is preferable to measure adiposity directly using such methods as bioelectrical impedance analysis (BIA) or dual-energy X-ray absorptiometry (DXA). To date, whether stunting is casually associated with later adiposity is unclear in view of the many method-

ologicall challenges that remain. For a discussion of possible pathways underlying the association between early life stunting and subsequent body composition and nutritional status, see Wells (2019).

Interpretive criteria

The distribution of height measurements at a given age within most populations is often narrow, so that accurate measuring techniques are essential. Moreover, a deficit in length takes some time to develop, so assessment of nutritional status based on length-for-age alone may underestimate malnutrition in infants in some settings.

Data for length-for-age for boys and girls age 0–6mos, 0–2y, 6mos–2y and height-for age 2–5y and length/height for age 0–5y expressed as Z-scores or percentiles based on the WHO Child Growth Standards are recommended for international use, as noted earlier. Both charts and tables are available (Growth Standards). The U.S. CDC 2000 growth charts also provide length-for-age from 0–36mos and stature-for-age 2–20y (24–240mo) (Kuczmarski et al., 2002).

Note the use of the new international standard for length at birth for gestational age developed in the INTERGROWTH-21st Project by Villar et al. (2014) will provide a method for the early diagnosis of linear growth failure, provided gestational age can be assessed correctly. The WHO Child Growth Standard can then be used to monitor linear growth failure during infancy and childhood.

10.2.5 Height-for-age difference

Height-for-age Z-scores (HAZ) are widely used to assess children's attained height at a given age (see Chapters 9 and 13 for more details on Z-scores) However, some investigators have used positive *changes* in attained mean height-for-age Z-scores to identify population-level catch-up growth in children (Crookston et al., 2010). Leroy et al. (2014) have raised concern over the appropriateness of using height-for-age Z-scores to evaluate such changes in linear growth with age over time because the cross-sectional standard deviations used in the denominator of height-for-age Z-scores and shown in Box 10.3 are constructed from cross sectional data and are not constant over time, but increase linearly from birth to 5 years of age. As a result, a child with a constant absolute height deficit will appear to improve with age based on the HAZ. Instead, Leroy et al. (2014)recommend using height-for-age difference (HAD) to describe and compare height changes as populations of children age. Height-for-age difference (in cm) is defined as: child's height compared to standard, expressed in centimeters. It is calculated by subtracting the sex- and age-specific median height (from the WHO Child Growth Standard) from the child's actual height as shown in Box 10.3.

Box 10.3 Height-for-age Z-scores (HAZ) and height-for-age difference (HAD)

- Growth deficits in height in groups of children are expressed as the mean of the individual deficits. These are calculated as the difference between the measured height and the median age- and sex-specific height from the 2006 WHO growth standard. This HAD can be used in absolute terms or it can be used standardized by dividing HAD by the SD from the growth standards to calculate HAZ:

$$\text{HAD} = \text{observed height} - \text{median height growth standard}$$
$$\text{HAZ} = \frac{\text{observed height} - \text{median height growth standard}}{\text{SD growth standard}}$$

  Thus:

$$\text{HAZ} = \frac{\text{HAD}}{\text{SD}}$$

- HAZ is constructed using cross-sectional SDs. HAZ is useful to assess the attained height of children at a given age but is inappropriate to assess changes in height as children age; HAZ is thus inappropriate to assess catch-up growth in height (Leroy et al., 2014). Assessing catch-up growth using HAZ is mathematically different from using HAD and has been demonstrated to lead to erroneous conclusions (Leroy et al., 2015).
  From Leroy et al. (2020)

Figure 10.20 compares changes in growth in populations of children between 0–60mos based on HAD and HAZ using data from 51 nationwide surveys from low- and middle-income countries (Leroy et al., 2014). Note in the Figure 10.20, the mean HAZ started below the WHO Child Growth Standard (at approximately –0.4 Z-scores) and fell markedly up to 24mos, after which it stabilized and increased only slightly. In contrast, based on the mean HAD curve, the children started with an average height deficit of 0.8cm, with the most pronounced faltering (i.e. steepest slope) evident between 6–18mos, as shown for the HAZ curve. Nevertheless, in contrast with the trend in the HAZ curve, the deficits in linear growth depicted by the HAD curves continued to increase after 18mos to 60mos, although at a lower rate than before 18mo. Indeed, the slopes of the HAD curves provide no indication that the process of growth faltering leveled off even at 5y. The bumps in the curves just after 24, 36, and 48mos are reportedly due to the tendency to report age in completed years rather than exact months (Victora et al., 2010). Such growth deficits were also evident in five regions (E. Europe and Central Asia, North Africa and Middle East, Latin America and Caribbean, Africa South of the Sahara, South Asia), although the magnitude of the deficits varied between regions; see Leroy et al. (2014) for more details. For more examples comparing changes in growth in populations of children age 2–5y using HAD vs. HAZ based on both cross-sectional and longitudinal data, see Leroy et al. (2015). For a discussion of whether linear growth retardation (or "catch-up growth") and the associated negative effects are reversible, consult Leroy et al. (2020).

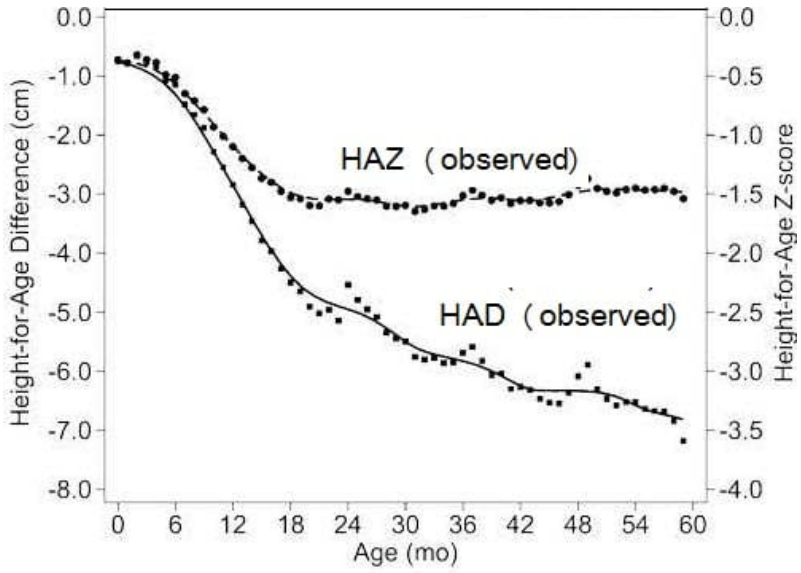


Figure 10.20. Mean HAZ and HAD relative to the WHO standard (1–59mo) by completed month and kernel-weighted local polynomial smoothed values. Data from n = 309,215 children from 51 demographic and health surveys and multiple indicator cluster surveys. HAD, height-for-age difference; HAZ, height-for-age Z-score. Redrawn from Leroy et al. (2014).

10.2.6 Classification of the severity of malnutrition based on the prevalence of wasting, overweight, and stunting

UNICEF/WHO/World Bank Group (2023) update the joint global and regional estimates of malnutrition among children under 5 years of age each year. These estimates of prevalence and numbers affected for child stunting, overweight, wasting and severe wasting are derived for the global population as well as by regional groupings of United Nations. A new country-level model was used to generate the country, regional and global estimates for stunting and overweight, for the 2021 edition and annual estimates from 2000 to 2020 are newly available. These estimates do not account for the impact of COVID-19, but the pandemic is expected to exacerbate all forms of malnutrition. This is likely due to worsening household income especially in vulnerable populations, constraints in the availability and affordability of nutritious food, disruptions in essential nutrition services, and reduced physical activity.

The severity of malnutrition in young children age < 5y was originally classified into five classes of severity based on the prevalence (as %) of wasting, stunting, and overweight. This practice was adopted to highlight the levels and trends across countries, and identify the areas of greatest need and hence likely to gain the most benefit from an intervention (de Onis et al., 2019). The five prevalence levels are shown in (Table 10.2).

Table 10.2 Prevalence thresholds, corresponding labels, and the number of countries (n) in different prevalence threshold categories for wasting, overweight and stunting in children under 5y using the "novel approach". Wasting: weight-for-length/height < –2 Z-score; Overweight: weight-for-length/height > +2 Z-score; Stunting: length/height-for-age < –2 Z-score; From: de Onis et al. (2019).

Wasting			overweight			Stunting		
Prevalence thresholds (%)	Labels	(n)	Prevalence thresholds (%)	Labels	(n)	Prevalence thresholds (%)	Labels	(n)
< 2·5	Very low	36	< 2·5	Very low	18	< 2·5	Very low	4
2·5 – < 5	Low	33	2·5 – < 5	Low	33	2·5 – < 10	Low	26
5 – < 10	Medium	39	5 – < 10	Medium	50	10 – < 20	Medium	30
10 – < 15	High	14	10 – < 15	High	18	20 – < 30	High	30
≥ 15	Very high	10	≥ 15	Very high	9	≥ 30	Very high	44

The fifth threshold labeled "very low" of no-public health concern was included across all three indicators to reflect the expected prevalence in a healthy population. Theoretically, the proportion of children with a Z-score < –2 or > +2 relative to the median of the WHO Child Growth Standards in a healthy population is 2.3% (rounded to 2.5%). The higher four prevalence thresholds are defined by multipliers of this "very low" level. As an example, Figure 10.21 depicts a modeled percentage of children under 5y affected by stunting in countries worldwide in 2020, categorized by the five prevalence thresholds for stunting presented in Table 10.2.

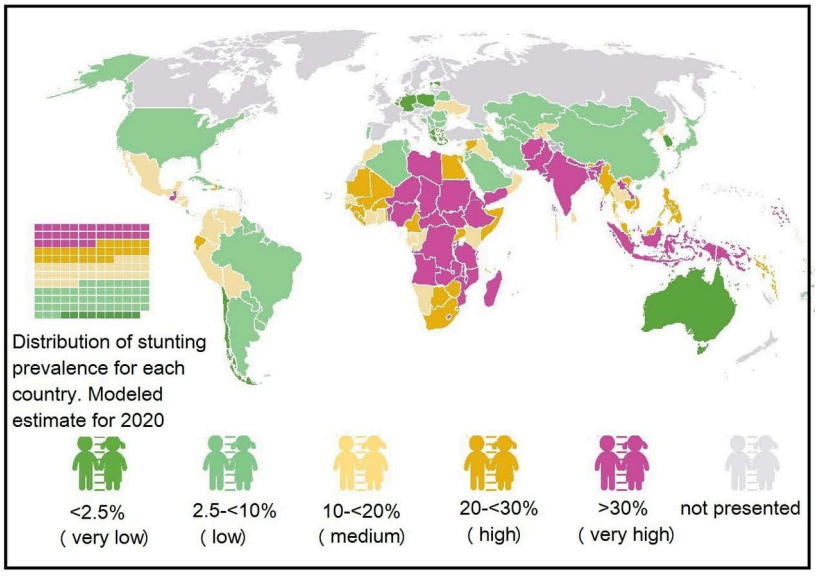


Figure 10.21. Modeled percentage of children < 5y affected by stunting in 2020. From: UNICEF, WHO, WB (2023)

10.2.7 Composite Index of Anthropometric Failure (CIAF)

Currently, the UNICEF/WHO/World-Bank classification scheme provides no estimates on children suffering simultaneously from multiple anthropometric deficits such as stunting plus wasting. However, there is increasing recognition that individual children may be at risk of both conditions simultaneously, might be born with both, pass from one state to the other over time, and accumulate risks to their health and life through their combined effects.

In response to these concerns, Nandy and Svedberg (2012) proposed the Composite Index of Anthropometric Failure (CIAF). This

is an aggregate measure aimed to estimate the overall burden of undernutrition in children age < 5y that incorporates children who are wasted and/or stunted and/or underweight based on the WHO Child Growth Standard (Growth Standards). The original CIAF model identifies six groups of children defined by the categories B to G, shown in Table 10.3. The overall CIAF excludes those children not in anthropometric failure (i.e., group A) and counts all children who have wasting, stunting, or are underweight (i.e., groups B to F), thus providing a single measure to estimate the overall prevalence of undernutrition (Porwal et al., 2021) (Table 10.3).

Table 10.3 Classification of anthropometric failure as per Composite Index of Anthropometric Failure (CIAF) categories, together with the distribution of participants per CIAF category. Note: the combination of wasted and stunted is not included as it is physically impossible because a child cannot have a low weight-for height and a low height-for-age and not be underweight. Data from from Porwal et al. (2021).

	Categories of Undernutrition	Wasting	Stunting	Under-weight	N = 30,500	(%)
A	No failure	No	No	No	18,434	51.8
B	Wasting (Low weight-for-height WTHT) only	Yes	No	No	1392	4.6
C	Wasting and Underweight [Low weight-for-height (WTHT) and low weight-forage (WTA)]	Yes	No	Yes	1729	6.5
D	Wasting, Stunting and Underweight (All three anthropometric failures)	Yes	Yes	Yes	1222	6.0
E	Stunting and Underweight [Low height-for-age (HTA) and Low weight-for-age (WTA)]	No	Yes	Yes	3552	16.6
F	Stunting only [Low height-for-age (HTA)]	No	Yes	No	3467	11.5
Y	Underweight only [Low weight-for-age (WTA)]	No	No	Yes	704	3.0
Overall CIAF	B + C + D+ E + F + Y				12,066	48.2

Anthropometric data comparing the percentage of Indian children classified as under-nourished when based on one of the three conventional growth indices and using the CIAF as an aggregate indicator of undernutrition are also shown in Table 10.3. The data are based on 38,060 Indian children aged < 5y collected in the 2016–2018 Comprehensive National Nutrition Survey (CNNS) (Porwal et al., 2021). As shown in Table 10.3, the prevalence rate (as %) of undernutrition is consistently less when only one of the three conventional indices is employed, compared to the overall CIAF prevalence of 48.2% among children < 5y. Of these Indian children, 19.1% had only one form of anthropometric failure (groups B, F and Y), whereas 22.1% had two forms of anthropometric failure simultaneously (groups C and E), and 6% had all three forms of anthropometric failures (Group D).

Nandy and Svedberg also argued that children who were simultaneously stunted, wasted, and underweight were likely to be at the greatest risk of mortality. However, most of the studies which have investigated risk of child mortality using the CIAF model to classify under-nutrition, have been cross-sectional (Khan & Das, 2020) so no causal inferences could be drawn. McDonald et al. (2013) were the first investigators to estimate mortality risk associated with multiple anthropometric deficits based on prospective cohort data and classified according to the CIAF model (Table 10.3). The data were from ten prospective cohorts in Africa, Asia, and Latin America. Prospective cohort designs, unlike cross-sectional studies, provide the best information about causation of disease and the most direct measurements of the risk of developing disease.

McDonald et al. (2013) from their meta-analysis, reported a significantly elevated risk of all-cause mortality among children with 1, 2 and 3 anthropometric deficits compared with children with no deficits, with a dose response relationship in relation to the number of deficits. Moreover, for children wasted and underweight but not stunted, risk of mortality was greater than for those who were stunted and underweight but not wasted. Cause-specific mortality, however, could not be investigated due to insufficient statistical power; see McDonald et al. (2013) for further details.

More studies are needed to establish whether there are unique determinants of multiple anthropometric deficits, as well as

to identify the combination of anthropometric deficits associated with the greatest risk of mortality. With this knowledge, interventions could be targeted to the subgroup of children who may benefit the most.

Table 10.4 Prevalence of malnutrition based on the CIAF among children aged 6–59mos, in the Ghana Socioeconomic Panel Survey (GSPS). N = 1608, calculated using the composite index of anthropometric failure (CIAF) and the extended composite index of anthropometric failure (eCIAF). Data from Kuwornu et al. (2020).

Group	Anthropometric Status	CIAF - Prevalence (95% CI)
A	No failure	54.5 (50.9–58.1)
B	Wasting only	4.2 (3.0–5.5)
C	Wasting and underweight	9.0 (7.2–10.8)
D	Stunting, wasting, and underweight	5.6 (4.3–7.0)
E	Stunting and underweight	7.6 (5.9–9.2)
F	Stunting only	18.4 (15.8–21.0)
Y	Underweight only	0.8 (0.3–1.2)
B to Y	Overall undernutrition (CIAF)	45.5 (42.0–49.1)
		eCIAF - Prevalence (95% CI)
A	No failure	42.7 (39.4–46.1)
B	Wasting only	4.2 (3.0–5.5)
C	Wasting and underweight	9.0 (7.2–10.8)
D	Stunting, wasting, and underweight	5.6 (4.3–7.0)
E	Stunting and underweight	7.6 (5.9–9.2)
F	Stunting only	12.3 (10.2–14.4)
Y	Underweight only	0.8 (0.3–1.2)
G	Stunting and overweight	6.1 (4.7–7.4)
H	Overweight only	11.7 (9.5–14.0)
B to H	Overall malnutrition (eCIAF)	57.3 (53.9–60.6)

With the emergence of the dual burden of malnutrition, the original CIA model of classification for under-nutrition is no longer sufficient. This has led to the extension of the CIAF model to include two new groups: “stunting and over-weight” and “overweight only”, as shown in Table 10.4. Overweight is defined as a BMI-for-age Z-score > +2 for children < 5y based on the WHO Child Growth Standard (de Onis & Lobstein, 2010). The prevalence of malnutrition based on the eCIAF among children aged 6–59mos who participated in the 2011 Ghana Socioeconomic Panel Survey (GSPS) (n=1608) is also presented in Table 10.4 (Kuwornu et al., 2020). From the estimate of eCIAF in Table 10.4, more than one half of all Ghanian children aged 6–59mos had at least one form of malnutrition over the study period.

10.2.8 Weight changes

Body weight is the sum of the protein, fat, water, and bone mass in the body. Changes in body weight do not provide any information on the relative changes among these components. Increasingly changes in the components of body composition are assessed using laboratory-based methods, such as dual-energy X-ray absorptiometry (DXA), deuterium dilution, or bioelectrical impedance analysis. For details of these techniques, see Chapter 14.

Body weight is the sum of the protein, fat, water, and bone mass in the body. Changes in body weight do not provide any information on the relative changes among these components. In normal adults, there is a tendency for increased fat deposition with age, concomitant with a reduction in muscle protein. Such changes are not evident in body weight measurements but can be seen by determining either body fat or the fat-free mass. In healthy persons, the daily variations in body weight are generally small

(i.e., less than ±0.5kg). In conditions of acute or chronic illness, however, negative energy-nitrogen balance may occur as the body can use endogenous sources of energy (including protein) as fuel for metabolic reactions. Consequently, body weight declines. In conditions of total starvation, the maximal weight loss is approximately 30% of initial body weight, at which point death occurs. In chronic semistarvation, body weight may decrease to approximately 50%–60% of ideal weight. In contrast, when persistent positive energy balance occurs, there is an accumulation of adipose tissue, and body weight increases.

Body weight can only be used to assess the severity of undernutrition in subjects with relatively uncomplicated, nonedematous forms of semistarvation (Heymsfield et al., 1984). In disease conditions in which edema, ascites (fluid in the abdominal cavity), dehydration, diuresis, massive tumor growth, and organomegaly occur, or in obese patients undergoing rapid weight loss, body weight is a poor measure of body energy-nitrogen reserves. In such conditions, a relative increase in total body water, for example, may mask actual weight loss that results from losses of fat or skeletal muscle. Massive tumor growth may also mask losses of fat and muscle tissue, which may occur during severe undernutrition. Hence, additional anthropometric measurements (e.g., mid-upper-arm circumference and triceps skinfold thickness) should be taken to obtain more information on the origin of any change in body weight (Heymsfield et al., 1984).

To assess weight changes, the actual and usual weight of an individual must be known. From these two measurements, the percentage of usual weight, percentage of weight loss (or weight gain), and rate of change can be calculated using the equations shown in Box 10.4. The individual's actual weight can also be compared with appropriate age- and sex-specific reference data.

Box 10.4

$$\%\ \text{Usual wt.} = \frac{\text{actual wt.}}{\text{usual wt.}} \times 100\%$$

$$\%\text{Wt. loss} = \frac{\text{usual wt.} - \text{actual wt.}}{\text{usual wt.}} \times 100\%$$

$$\text{Rate of change} = \frac{\text{body weight present} - \text{body weight initial}}{\text{day present} - \text{day initial}} (\text{kg/d})$$

Weight changes in children

In low income countries, where there is a high prevalence of under-nutrition, weekly weight gain has often been used to monitor short-term response of children to a feeding program. For example in the past, in children with severe acute malnutrition (SAM), percentage weight gain was applied as the discharge criterion. However, use of a weight-based assessment can be misleading in children with SAM. Such children may have diarrheal disease accompanied by dehydration or edema, both of which have an affect on body weight (Modi et al., 2015).

Both lack of weight gain and weight loss in a child have been shown to be independently and more closely related to mortality than other indicators of undernutrition, such as BMI-for-age based on survival data from 2402 rural children aged 0–24mos fom the Democratic Republic of Congo (O'Neill et al., 2012). These two indicators are also included in a set of diagnostic indicators developed to identify pediatric under nutrition by the the U.S. Academy of Nutrition and Dietetics and the American Society for Parenteral and Enteral Nutrition (ASPN) (Becker et al., 2015). Pediatric undernutrition in developed countries generally occurs in a setting of acute or chronic illness. Nevertheless more studies are needed to evaluate the feasibility, suitability, validity, and reliability of these indicators to identify pediatric undernutrition (Mogensen et al., 2019).

With the growing concern about the dual burden of nutrition, changes in body weight are being monitored in apparently healthy children. For example, changes in body weight over a 15y period (both weight loss and weight gain) have been monitored in apparently healthy children and adolescents (n=> 150,000) in Germany (Vogel et al., 2022). A change outside ±0.2 BMI-SDs per year was classified as a high positive or high negative change. Between 2005–2020, there was a small but stable positive trend in the proportion of children with a high positive weight change, especially within those children with obesity. This trend was accompanied by a decrease in the proportion with a high negative weight change. During COVID-19, the weight gain was substantial across all weight and age groups (i.e., 1–6y; 6–12y; 12–18y), as noted by others (Jia, 2021; Brooks et al., 2021): COVID-19 has aggravated the childhood obesity pandemic.

Weight changes in adults

The U.S Academy of Nutrition and Dietetics and the American Society for Parenteral and Enteral Nutrition (ASPN) have also devised a series of diagnostic characteristics to identify and document malnutrition (undernutrition) in adults in acute, chronic,

and transitional care settings in developed countries. Six clinical charateristics were identified to support a a diagnosis of malnutrition in adults, one of which was percentage weight loss from baseline (White et al., 2012). For more details of these diagnostic indicators, the reader is referred to Chapter 27, Nutritional assessment of hospital patients.

Marked changes in body weight in healthy adults with age have been documented in several national surveys in industrialized countries. Such changes may have long-term consequences on health. In general during the period from young to middle adulthood, adults gain weight more rapidly and accrue excess adiposity, whereas from middle to late adulthood, weight begins to stabilize or even decrease, especially around age 70y, when the decrease may continue for the remainder of life.

Several investigators have examined the association between weight changes across adulthood and mortality. For example, in a prospective cohort study of US adults participating in NHANES from 1988 to 2014, Chen et al. (2019) reported stable obesity across adulthood, weight gain from young (age 25y) to middle adulthood (mean age 47y), and weight loss from middle to late adulthood, were associated with increased risk of mortality. These findings suggest that maintaining a normal healthy weight across adulthood, and especially preventing weight gain in early adulthood, is important for preventing premature deaths in later life.

The potential effect of weight loss and weight gain among older adults has also been investigated. In a systematic review and meta-analysis of community-dwelling older adults aged > 65y, the effect of weight loss on increased risk of all-cause mortality was stronger than weight gain (Alharbi et al., 2021), as noted by others (Karahalios et al., 2017). However, the point at which weight loss may indicate increased mortality risk has not been clearly defined. Further research is needed to determine whether these associations vary with gender, initial weight, and whether or not the weight loss/gain was intentional.

Nevertheless, caution when interpreting the results of some of these epidemiological studies is warrented. Methodological issues of concern include confounding by smoking, reverse causation due to existing chronic diseases, especially among the elderly, and non-specific loss of lean mass and function in the frail elderly (Fontana & Hu, 2014).

Gestational weight changes

The U.S. National Research Council (2009) have compiled new guidelines for gestational weight gain in response to the increase in the proportion of U.S. women of reproductive age who are overweight and obese. Gestational weight gain includes gains in maternal and fetal fat-mass and fat-free mass, as well as the placenta and amniotic fluid. How these different components of gestational weight gain influence both maternal and offspring health remains uncertain: see Widen and Gallagher (2014) for more details.

For the first time, the new U.S guidelines consider those health outcomes of both mother and child during and after delivery which are plausibly related to gestational weight gain, and the trade-offs between them. For the mother, the most important health outcomes considered were postpartum weight retention and unscheduled cesarean delivery, whereas for the infants the outcomes were preterm birth, extremes of birthweight (expressed as small- or large-for gestational age), and childhood obesity (Rasmussen et al., 2010).

Table 10.5 presents the new recommendations for total weight gain ranges for U.S women according to pre-pregnancy BMI (kg/m^2) based on the BMI categories of WHO. The rates of weight gain (mean and range) for the 2nd and 3rd trimester are also stated. These recommendations have been adopted by WHO for low-income populations, until specific country cutoff values are available. Note the recommendations that were included in the earlier IOM (1990) guidelines for short, young, or black women are no longer included in the new U.S. IOM/NRC (2009) recommendations.

*Table 10.5. New recommendations for total and rate of weight gain during pregnancy, by prepregnancy BMI. * Calculations assume a 0.5–2kg (1.1–4.4lbs) weight gain in the first trimester(Rasmussen et al., 2010 and based on Siega-Riz et al., 1994; Abrams et al., 1995; Carmichael et al., 1997).*

	Total Weight Gain		Rates of Weight Gain* 2nd and 3rd Trimester	
Prepregnancy BMI	Range in kg	Range in lbs	Mean (range) in kg/week	Mean (range) in lbs/week
Underweight (< 18.5 kg/m^2)	12.5–18	28–40	0.51 (0.44–0.58)	1 (1–1.3)
Normal weight (18.5–24.9 kg/m^2)	11.5–16	25–35	0.42 (0.35–0.50)	1 (0.8–1)

	Total Weight Gain		Rates of Weight Gain* 2nd and 3rd Trimester	
Prepregnancy BMI	Range in kg	Range in lbs	Mean (range) in kg/week	Mean (range) in lbs/week
Overweight (25.0–29.9 kg/m^2)	7–11.5	15–25	0.28 (0.23–0.33)	0.6 (0.5–0.7)
Obese (≥ 30.0 kg/m^2)	5–9	11–20	0.22 (0.17–0.27)	0.5 (0.4–0.6)

The new U.S. recommendations include a relatively narrow range of recommended weight gain for obese women. However, because of the lack of data, no recommendations were made to reduce the guidelines for obese women to below 5.0–9.0kg, or specific guidelines according to obesity classes II and III. Women with a pre-pregnancy BMI indicative of obesity are recommended to lose weight before and not during pregnancy, and clinicians are encouraged to monitor weight gain at each prenatal visit. Interventions may be needed to assist women, especially those who are overweight or obese at the time of conception, to meet these guidelines.

This page titled 9.3: Growth indices, indicators, and recommended growth reference data (10.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

9.4: Body mass index in adults (10.3)

Weight to height ratios indicate body weight in relation to height and are particularly useful for providing a measure of overweight and obesity in adult populations. Hence these ratios are sometimes referred to as obesity indices.

At the present time, the ratio most used in this way is the body mass index (BMI) (also termed Quetelet's index), calculated as weight (kg) / height^2 (m). Body mass index is used in preference to other weight / height indices, including the weight / height ratio, the Ponderal index, and Benn's index (Table 10.6). BMI is now widely used to classify overweight and obesity in adults.

Table 10.6. Indices for weight relative to height. The power p in Benn's index is calculated to minimize the direct relationship with height (Benn, 1971).

Index	Formula
Weight/height ratio	wt/ht
Body Mass Index	wt/(ht)^2
Ponderal index	ht / ∛wt
Benn's index	wt/(ht)^p

Body mass index is relatively unbiased by height and correlates reasonably well with other anthropometric measures of adiposity in adults such as waist circumference (WC) and abdominal visceral fat (AVF), as shown in Figure 10.22. For details on the importance of waist circumference and abdominal visceral fat, see Chapter 11.

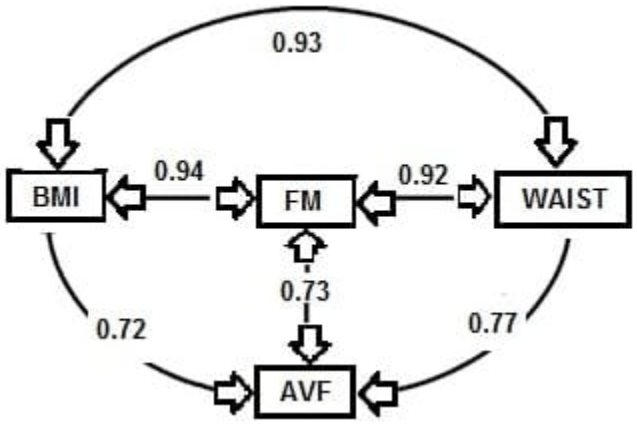


Figure 10.22. Correlations among BMI, FM, WAIST and CT-assessed AVF. The arrows are bidirectional to indicate that no causal relationships are implied in the model. Each coefficient represents the weighted mean of six correlations computed in samples of black men, black women, white men and white women from the HERITAGE Family Study, and of Caucasian men and women from the Quebec Family Study. AVF, Abdominal visceral fat; FM, fat mass; waist, waist circumference. From Bouchard (2007).

Body mass index is employed in large-scale nutrition surveys and epidemiological studies as a measure of overweight and obesity; measurements of weight and height are easy, quick, relatively noninvasive, and more precise than skinfold thickness measurements. Nevertheless, BMI does not quantify total body adiposity or distinguish between weight associated with muscle and weight associated with body fat. Hence, in some circumstances, an elevated BMI may result from excessive adiposity, muscularity, or edema, and is thus an inaccurate measurement for adiposity at the individual level.

Certain population groups may also be misclassified as overweight or obese based on BMI, due to differences in body proportions and the relationship between BMI and body fat content. Māori and some Polynesian groups, for example, appear to have less body fat than Caucasians of the same age, sex, and BMI (Craig et al., 2003; Rush et al., 2009), but in some Asian populations, the reverse is true(Deurenberg et al., 1998; WHO Expert Consultation, 2004) (Section 10.3.1). Other examples include

athletic populations and individuals who are very short (< 150cm). Abnormal relationships between leg and trunk length can also lead to some misclassification (Deurenberg et al., 1999), although BMI can now be corrected to consider unusual leg length based on the ratio of sitting height to standing height(Norgan & Jones, 1995). Such adjustments are necessary for example, among Australian Aboriginals.

In addition, BMI conveys no information about the actual distribution of body fat. Anomalies in the distribution of abdominal visceral fat, and in particular ectopic fat (i.e., fat depots in liver, heart, muscle, pancreas and kidney) are now recognized to be an even greater cardiometabolic risk than excess body fat *per se* (Neeland et al., 2019). Consequently, BMI alone can misrepresent the burden of abdominal visceral fat in a subgroup of individuals who are at highest cardiometabolic risk. Therefore, a BMI within the normal range does not necessarily preclude an increased cardiometabolic risk. In a community-dwelling adult men in Korea, for example, assessment of cardiometabolic risk based on high blood pressure, hyperglycemia and dyslipidemia, was highest among those with a BMI < 25kg/m^2 but who had a high waist circumference (i.e., > 90cm). Moreover, in this group the amounts of both total fat mass and percentage body fat (assessed via bioelectrical impedance analysis) were similar to the group with a BMI > 25kg/m^2 and a waist circumference < 90cm (Kim et al., 2020).

Therefore, increasingly a combination of both waist circumference (as a surrogate estimate of abdominal visceral fat) and BMI are being used to predict risk of cardiovascular disease (Nazare et al., 2015). In the future, simple clinically applicable tools to monitor changes over time in abdominal visceral fat and ectopic fat need to be developed for use with BMI and waist circumference. Currently, the laboratory methods available are used mainly in research settings; they are described in Chapter 14.

10.3.1 BMI and measures of body fat and disease risk

The validity of BMI as an index of the percentage body fat and fat mass (in kg) in adults has been assessed by comparing BMI with body fatness estimated initially using a 2-component model, although more recently multi-component models of body composition are being used (Gallagher et al., 1996; Wells et al., 2010; Silva et al., 2013).

In the simpler 2-component model, total body mass is partitioned into total body fat and fat-free mass, applying the principle that if one of the components is measured, the other can be estimated. However, when using this approach, the calculation of body fat depends on certain theoretical assumptions. The hydration of fat-free mass, for example, is assumed to be constant within and between individuals even though inter-individual variability is known to exist, especially during growth and maturation in children, in pregnancy, and among adults with varying adiposity (Wells et al., 2010; Most et al., 2018; Gutiérrez-Marin et al., 2021; Gallagher et al., 1996).

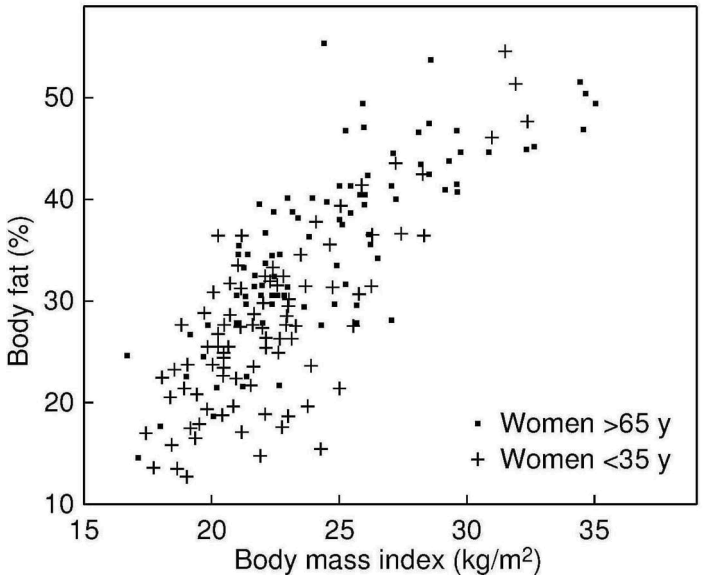


Figure 10.23. The relationship between body mass index and percentage of body fat in both younger and older women from New York City. Body fat determinations based on a four-compartment bodycomposition model. Data from Gallagher et al., (1996).

In contrast, use of multi-component models for measuring body composition minimizes the use of theoretical assumptions for key body properties. In Figure 10.23, for example, the relation between BMI (kg/m^2) and body fat (as a percentage) was assessed using a 4-component model in young (< 35y) and elderly (> 65y) women. A comparable relation between BMI and body fat mass (kg) was also observed. In this study, measurements were collected on body weight, body volume, total body water, and bone mineral mass, allowing the inter-individual variability in the composition of the fat-free mass to be considered rather than assuming fixed constants for the water content, bone mineral content, or density of fat-free mass (Wells & Fewtrell, 2006).Consequently, the four-component model is the established criterion reference method for generating accurate data on body fat and fat-free mass. For more details of these methods used to assess body composition, see Chapter 14

Over the last few decades, validation studies have analyzed the performance of BMI to detect body adiposity by comparison with a variety of techniques considered to measure body composition accurately. They have included a combination of statistical approaches including correlation and regression analyses, paired t-tests, and more recently, Bland-Altman analyses. In some cases, the sensitivity and specificity of the methods have also been compared (Okorodudu et al., 2010). For details of the statistical techniques now applied to evaluate agreement between body composition methods used in validation studies, see Earthman (2015).

A major assumption of BMI has been that it is an independent index of body fat. This means that after adjusting for body weight-for-stature, all subjects with the same BMI have the same relative fatness, irrespective of their age, sex, or ethnicity. However, it is now widely known that the relationship between BMI and percentage body fat is influenced by age, sex, and ethnicity, all of which have implications for the use of BMI as an index of body fatness (Norgan, 1994; Gallagher et al., 1996; Jackson et al., 2002).

It is recognized that the relationship between BMI and body fat is both age- and sex-dependent (Deurenberg et al., 1991; Gallagher et al., 1996). For example, in the cross-sectional study shown in Figure 10.23, older women tended to have a relatively greater percentage of body fat than their younger counterparts with a comparable BMI (Gallagher et al., 1996). A similar age-related trend was noted for men; these age-related trends persist up to 60–65y of age in both sexes. According to a recent critical review based on data from the U.S NHANES survey, such age-related changes in body composition by both men and women possibly reflect senescence-mediated hormonal changes, lowering of physical activity levels, reductions in trunk length owing to osteoporosis, and a variety of other related mechanisms (Heymsfield et al., 2016). Moreover, women have significantly greater amounts of total body fat than do men for an equivalent BMI suggesting that BMI cannot be used as a comparable measure of fatness in males and females. These sex differences are substantial and are maintained throughout the entire adult life span (Gallagher et al., 1996). Further, they can be seen in children as young as 3–8y (Taylor et al., 1997).

The relationship between BMI and the percentage of body fat appears to differ among certain race/ethnic groups, as noted earlier. Figure 10.24 shows results of a meta-analysis among different race/ethnic groups conducted by Deurenberg et al. (1998).

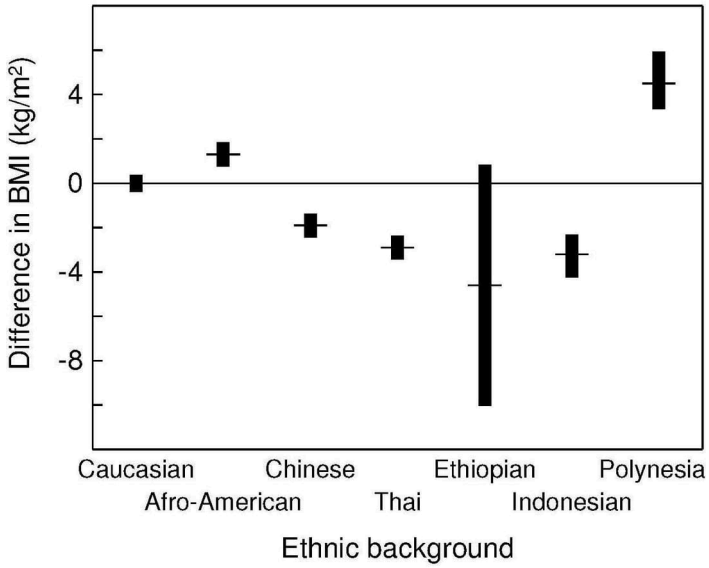


Figure 10.24. Ethnic differences in calculated values for body mass index (BMI) (mean and 95% confidence interval) which reflect equal levels of body fat, adjusted for age and sex. The means are relative to the results for Caucasians (set to 0.0). Data from Deurenberg et al. (1998).

Of the 15 studies included in the meta-analyses, 11 used those considered by the investigators to be a "reference criterion method" for the measurement of body fat, and included hydrostatic weighing, deuterium dilution, and DXA. Results of the meta-analysis indicate that in some populations (e.g., some Chinese, urban Thais, Indonesians), levels of obesity in terms of percentage of body fat will be greater at the 30kg/m^2 obesity cutoff suggested by WHO (2000) than for Europeans.

The existence of such race/ethnic-related differences has been confirmed by several other investigators (Wagner & Heyward, 2000; Deurenberg-Yap et al., 2000; Hsu et al., 2012). For example, Asian Indians are also reported to have more body fat than Europeans at any given BMI (WHO Expert Consultation 2004; Low et al., 2009),whereas for Chinese from Beijing and rural Thais, values are like those of Europeans (He et al., 2015). In contrast, in some Pacific populations, the percentage of body fat at a given BMI is lower (see Polynesians in Figure 10.13) (Swinburn et al., 1999). Possible reasons for these race / ethnic-related discrepancies may include differences in body shape and body composition, some of which may be inherited and others due to environmental and lifestyle factors. There is an urgent need for more in-depth studies of race / ethnic differences in body shape and composition and how these differences relate to clinically meaningful risks; see Heymsfield et al. (2016) for more discussion.

Body mass index has been used in numerous international population studies to assess disease risk among adults. Increasing BMI is clearly associated with a higher risk of high blood pressure, type 2 diabetes mellitus, cardiovascular disease (coronary heart disease and stroke) and cancer (WHO Expert Consultation 2004; Abdullah et al., 2010;Ni Mhurchu et al., 2004 ; Whitlock et al., 2009; Wiseman, 2008). Indeed, the relative risk for cardiovascular disease risk factors and cardiovascular disease incidence increases in a graded fashion with increasing BMI in all population groups. Nevertheless, the risk of type 2 diabetes appears to vary according to race/ethnicity and presence or absence of other traditional risk factors, with Chinese Americans, Hispanics, African Americans over a 10y period having a similar risk at lower BMI values compared with those for white participants (Rodriguez et al., 2021). These findings are based on data from a large prospective cohort analysis from the Multi-Ethnic Study of Atherosclerosis of US adults aged > 45y who were free of diabetes at baseline and in whom incident type 2 diabetes was defined as fasting glucose > 7.0mmol/L, or use of any diabetes medications.

Overweight and obesity as defined by BMI has also been associated with increased mortality based on both large scale prospective studies in single countries (Jee et al., 2006; Patel et al., 2014) and systematic reviews and meta-analysis in several regions or continents For example, Aune et al.(2016) conducted a systematic review and meta-analysis of 230 cohort studies of BMI and the risk of all-cause mortality in five regions (North America, South America, Europe, Australia, and the Pacific). Overweight and obesity were associated with increased risk of all-cause mortality, consistent with the findings of Di-Angelantonia et al. (2016), from a study across four continents (Europe, North America, east Asia and Australia and New Zealand). In the systematic review and meta-analysis of Aune et al. (2016), the lowest mortality was observed in those participants with a BMI of about 25.0

In addition, associations between BMI and musculoskeletal disorders, impairments in respiratory and physical functioning, quality of life, and mental illness such as clinical depression, anxiety, and other mental disorders have been reported by the (CDC)

10.3.2 WHO classification of overweight and obesity in adults

The WHO International Obesity Task Force defined obesity as a condition in which abnormal or excessive fat accumulation may impair health. (WHO, 2000). They have recommended the use of a graded classification of both overweight and obesity to:

- Compare weight status within and between populations
- Identify individuals and groups who are at increased risk of morbidity and mortality
- Identify priorities for intervention at both the individual and community levels
- Allow interventions to be evaluated appropriately

Table 10.7. Classification of overweight in adults according to body mass index (BMI). From (WHO, 2024) and (CDC, 2024)

Classification	BMI (kg/m^2)	Risk of comorbidities
Underweight	< 18.50	Low (but risk of other clinical prob
Normal weight	> 18.5 to < 25.0	Average
Overweight	25 to < 30.0	Increased
Class 1 obesity	30 to < 35.0	Moderate
Class 2 obesity	35 to < 40.0	Severe
Class 3 obesity	> 40.0	Very severe

Table 10.7 shows the classification of adults according to BMI recommended by WHO (2024). The classification is based primarily on the association between BMI and mortality. WHO considers BMI as the most useful population-level measure of overweight and obesity as it is the same for both male and female adults and for all ages of adults. This classification system is not intended for use with pregnant and lactating women, or persons < 18y.

WHO (1995a) have also suggested that a universal cutoff point of BMI ≥ 30 could be used to estimate and compare the prevalence of obesity within and between populations. Figure 10.25 shows the trends in BMI for adults in some selected regions in 1980, 2008, and 2016 by sex. The figure emphasizes the wide variation in prevalence across regions: the low prevalence in South Asia is particularly marked. The increase in obesity from 1980 to 2016 is clearly a world-wide phenomena and independent of sex and region (NCD-RiskC, 2017).

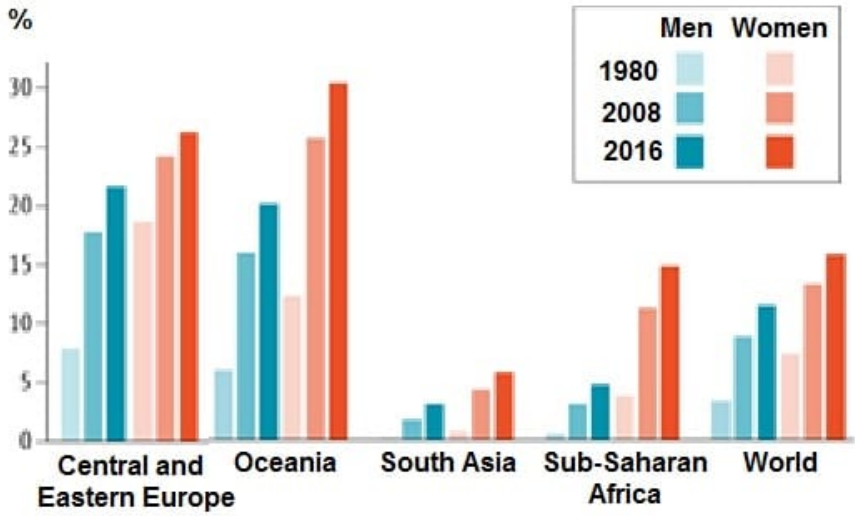


Figure 10.25. Global trends in the prevalence of obesity defined as BMI ≥ 30kg/m^2 in men and women in 1980, 2008 and 2016 from selected regions of the world. Data from the Noncommunicable Diseases Risk Factor Collaboration (NCD-RisC).

WHO (2000, 2004)recognized that these BMI cutoff points for overweight and obesity may not correspond to the same d.egree of fatness and risk of obesity-related co-morbidities across different race/ethnic populations, as noted earlier. In particular, in their 2004 report, WHO emphasized the mounting evidence that certain Asian populations have a particularly high risk of type 2 diabetes, cardiovascular disease, and mortality from other causes at relatively lower BMIs. However, WHO did not consider there was enough evidence for a single BMI cutoff point for all Asians for overweight or obesity at that time. Nevertheless, a report cosponsored by the WHO Western Pacific Region (WPRO) and led by the International Association for the Study of Obesity and the International Obesity Task Force (WHO/IASO/IOTF, 2000) did propose lower cutoff points for all Asian adults for overweight and obesity, based primarily on the relationship between BMI and morbidity; these are presented in Table 10.8. The use of these lower cutoffs for both overweight and obesity for Asians was supported by the findings of a prospective cohort study in Taiwan. (Wen et al., 2004).

Table 10.8. Proposed classification of weight by BMI for adult Asians.(WHO/IASO/IOTF, 2000).

Classification	BMI (kg/m^2)	Risk of comorbidities
Underweight	< 18.50	Low (but risk of other clinical problems is increased)
Normal weight	> 18.5 to 22.9	Average
Overweight	≥ 23	
At risk	23 to 24.9	Increased
Class 1 obesity	25 to 29.9	Moderate
Class 2 obesity	≥ 30.0	Severe

More recent research findings in South Asian populations suggest that a BMI ≥ 23kg/m^{2}is not sensitive enough to identify the obesity-associated disease risk in these populations, and indicate that an even lower cutoff (i.e., BMI ≥ 21kg/m^2) may be more appropriate in South Asian countries (Jauawardena et al., 2013).

10.3.3 Canadian classification of overweight and obesity in adults

The BMI classification system for adults used by Health Canada has been adapted from that of WHO (2000) The classification applies to adult women (nonpregnant and nonlactating) and men between the ages of 18 and 65y (Douketis et al., 2005).

Table 10.9. Health Canada (2003) classification of adults according to BMI. For persons ≥ 65y, the "normal" range may begin slightl into the "overweight" range.

Health Canada Classification	BMI (kg/m^2)	Risk of developing health probler
Underweight	< 18.50	Increased
Normal weight	18.50 to 24.99	Least
Overweight	25.00 to 29.99	Increased
Obese Class I	30.00 to 34.99	High
Obese Class II	35.00 to 39.99	Very high
Class 2 obesity	≥ 40.00	Extreme high

Health Canada (2003) cautions that the classification system may misclassify certain population groups such as adults with a very lean body build or who are highly muscular, young adults who have not reached full growth, those over 65y, and certain ethnic and racial groups. Several health problems associated with overweight and obesity are also specified. They include coronary heart disease, type 2 diabetes, hypertension, dyslipidemia, gallbladder disease, obstructive sleep apnea, and certain cancers. A combination of BMI with waist circumference as a measure of abdominal visceral fat are now used to screen for overweight and obesity and to assess a person's risk of health problems (i.e., Type 2 diabetes, hypertension, dyslipidemia and the metabolic syndrome, which combines three of these conditions) (Douketis et al., 2005).

10.3.4 U.S. classification of overweight and obesity in adults

The cutoff points that define normal weight, overweight, and the various categories of obesity in the United States are now the same as those used by WHO (2024) and are presented in Table 10.7.

Obesity is often subdivided into three categories in the United States, with class III (i.e., BMI > 40) designated as "severe obesity". These BMI cutoff points for adults define overweight and class I to III obese individuals and identify those who may be at increased risk for cardiovascular disease and other obesity-related conditions. However, within these categories, personal additional risk assessment is also needed because degree of risk can vary. For more details on the assessment and treatment of cardiovascular risk factors and obesity-comorbidities for individuals with overweight or class I to III obesity, the reader is referred to the AHA/ACC/TOS guideline (2013). Note that these guidelines also recommend the measurement of waist circumference in patients with BMI 25–34.9kg/m^2. The waist circumference cutoff points indicative of increased cardiometabolic risk used in the United States are > 88cm for women and > 102cm for men. These same waist circumference cutoffs are also used for adult Canadians. See Chapter 11 for more details of the measurement of waist circumference.

10.3.5 BMI and chronic energy deficiency in adults

In low-income countries, low values for BMI in adults have been consistently associated with a decline in work output, productivity, and income generating ability, as well as a compromised ability to respond to stressful conditions (Ferro-Luzzi et al., 1992).

Underweight has also been associated with increased mortality. For example, in an early study in India, a progressive increase in mortality rate below a BMI of 18.5 was reported, with an almost threefold higher rate after ten years for those with BMI below 16 (Naidu & Rao, 1994). Whether the individuals were ill prior to the measurement is unknown, however, making it difficult to assign the trend to any causal relationship. Nevertheless, some evidence for such a relationship has been reported based on more recent meta-analysis of prospective studies across continents (Di Angelantonio et al., 2016; Wild & Byrne, 2016), as well as from an analysis of the Global Burden of Disease Study (2017).

In developed countries, some of the early studies of low BMI and mortality failed to control for two principal sources of bias: cigarette smoking, and the elimination of thin adults with early mortality, many of whom may had been already ill when measured (Manson et al., 1987). More recently, Flegal et al. (2005) investigated excess deaths associated with underweight (as well as overweight and obesity) in the U.S NHANES data. In their study, they noted an increase in mortality with both underweight (BMI < 18.5) and with obesity (BMI ≥ 30), but not with overweight (i.e., BMI ≥ >25 but < 30). In most developed countries today, underweight is common in adults in hospitals and nursing homes, almost invariably secondary to disease, although in some more affluent countries underweight has been associated with osteoporosis, infertility, and impaired immunocompetence (Health

Canada, 2003).

A single cutoff point of < 18.5kg/m^2 is frequently used to define low BMI values, indicative of underweight in both men and women in both low-income and more affluent countries (Naidu & Rao, 1994; Health Canada, 2003; NCD Risk Factor Collaboration, 2017). FAO has categorized three degrees of underweight based on BMI and adopted the term "chronic energy deficiency" for underweight (Shetty & James, 1994). The FAO classification of adult chronic energy deficiency based on the three degrees of underweight are shown in Table 10.10.

Table 10.10. Simple classification of adult chronic energy deficiency using BMI (kg/m^2). From Shetty & James (1994).

Chronic energy deficiency grade	BMI (kg/m^2)
Normal	> 18.50
Grade I	17.0 to 18.4
Grade II	16.0 to 16.9
Grade III	< 16.0

WHO (1995a; 2004) adopted these same cutoffs to define three grades of low BMI, referred to as "thinness" in WHO (1995a) and as "underweight" in WHO (2004), rather than the FAO definition of "chronic energy deficiency." In WHO (2004), Grade I (17.0–18.4kg/m^2) is termed "mild underweight," Grade II (16.0–16.9kg/m^2) "moderate underweight, and Grade III (< 16kg/m^2) "severe underweight". WHO (1995a) also devised a classification system to identify populations with a public health problem on the basis of low BMI; details are shown in Table 10.11. However, challenges still exist in relation to the use of low BMI cutoffs for defining thinness or chronic energy deficiency in adult populations. It is likely that the cutoff for underweight, like overweight, might also differ with ethnicity and environment, as well as sex and duration of undernutrition, and more research is required.

Table 10.11. WHO classification of populations on the basis of low BMI. Compiled from WHO (1995a).

Situation	Percent of population with BMI (kg/m^2) < 18.
Low prevalence (warning sign, monitoring required)	5 to 9
Medium prevalence (poor situation)	10 to 19
High prevalence (critical situation)	20 to 39
Very high prevalence (critical situation)	≥ 40

In any healthy adult population, 3%–5% can be expected to have a BMI < 18.5kg/m^2, but in countries with food insecurity, an excessive proportion of the population may have BMIs < 18.5, with very few obese subjects. However, in general as socioeconomic conditions improve, there is a tendency for a population to shift from thinness to overweight (de Onis & Habicht, 1996). In many low- and middle-income countries, this trend has now occurred with a rising prevalence of overweight and obesity (Figure 10.25), which is surpassing the fall in the global prevalence of underweight (NCD Risk Factor Collaboration, 2017). For example, in 2014, more men were obese than underweight in 136 (68%) of countries based on a pooled analysis of trends on adult BMI from 1975–2014 from 200 countries and territories. For women, obesity surpassed underweight in 165 (83%) of countries and severe obesity surpassed underweight in 135 countries.

This emerging situation in which low and middle-income countries are impacted by both under- and overnutrition is termed the "double burden of malnutrition" (DBM) and is receiving increasing attention (Wells et al., 2020; Hawkes et al., 2020; Nugent et al., 2020). The WHO refers to the double burden of malnutrition as "characterized by the coexistence of undernutrition along with overweight, obesity or diet-related noncommunicable diseases (NCDs), within individuals, households and populations, and across the life-course" (WHO, 2017). This broad definition describes the current state of global health where rising rates of overweight and obesity overlap with stagnant or slowly decreasing rates of underweight and chronic energy deficiency.

This page titled 9.4: Body mass index in adults (10.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

9.5: BMI in children and adolescents (10.4)

Body mass index is also now the internationally recommended screening indicator of overweight and obesity in both children and adolescents (WHO, 1995a). The recommendation arises from the following observations:

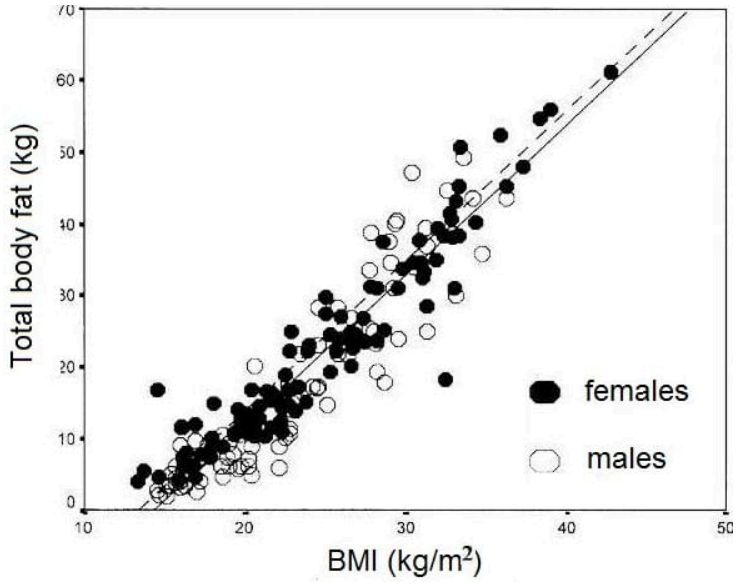


Figure 10.26 Relationship between BMI and total body fat (via DXA) in 90 boys and 98 girls aged 5–19y. Regression lines for boys (Model A:solid line) and girls (Model C;dashed line). Modified from Pietrobelli et al. (1998).

- Strong positive relationships between BMI and body fatness (measured in children by densitometry, magnetic resonance imaging, or DXA). Figure 10.26 presents the relationship between BMI and total body fat (via DXA) in 198 healthy Italian children and adolescents aged 5–19y
- Higher BMI among children in longitudinal studies is associated with tracking in the higher levels of blood pressure, serum lipids, and other factors that in adults are related to higher cardiovascular risk (Freedman & Sherry, 2009; Reilly et al., 2010; Kelly et al., 2013).

Notwithstanding the relation in some children of BMI and body fatness measured by DXA, BMI is not a perfect indicator of body fatness. Dietz and Bellizzi (1999) have cautioned that BMI will falsely classify some children of normal fatness as overweight and some overweight children as not overweight. Although errors in the measurement of percentage of body fat may account for some of the variability in the relationship between BMI and body fatness in children, several other factors may be implicated. These include enhanced muscular development, large head size, and a high torso-to-leg ratio. All these factors may falsely elevate BMI into an overweight range in some “non-overweight” children (Roberts & Dallal, 2001).

Defining overweight and obesity in children is more difficult than in adults. In children BMI has a characteristic curvilinear shape with age so that no simple cutoff can be used to define overweight or obesity, so the age-independent adult BMI classifications are not appropriate. Indeed, the 50th percentile for BMI for younger children and adolescents increases markedly from birth to early adulthood. These challenges have led many countries to use their national reference to compile their own ways of defining overweight and obesity in children making it impossible to make comparisons across countries (Rolland-Cachera, 2011).

There are three main classification schemes that are used internationally to define overweight and obesity in children. They are all based on BMI percentiles or SD scores derived from growth charts to account for gains in weight relative to height and each is described below.

10.4.1 International Obesity Task Force (IOTF) mass index cutoffs for overweight and obesity in children

In 1998, WHO convened an International Task Force on Obesity to define overweight and obesity in children and adolescents. This expert group agreed to recommend the adoption of an earlier approach developed by the European Childhood Obesity group (Poskitt, 1995) but based on an international reference population.

To accomplish this recommendation, Cole et al. (2000) compiled data for BMI for children 2–18y from six large nationally representative cross-sectional growth studies (Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States). The data were used to construct two country-specific percentile curves passing through the adult definitions of both overweight (BMI > 25kg/m^2) and obesity (> 30kg/m^2) at age 18y for each sex. The curves were then averaged across countries by age to yield sex-specific curves for each cutoff.

A major disadvantage of these international (IOTF) cutoffs, however, was that they could not be expressed as percentiles (e.g., 85th percentile) or standard deviation scores (+2 Zscores). As a result they could not be compared with other reference data in which the BMI cutoffs were defined by age-sex- specific BMI percentiles (e.g., 85th or 95th percentile) or standard deviation scores (e.g., +2SDs). Consequently, Cole and Lobstein (2012) reformulated their original international cutoffs using the LMS curves which permit BMI to be expressed as a percentile or SD score. Table 10.12 presents the sex-specific BMI cut- offs at age 18y for overweight (BMI 25), obesity (BMI 30), and morbid obesity (BMI 35), expressed as a SD score equivalent or percentile equivalent. Only small differences (< 0.2% on average) were observed when these new international IOFT cutoffs were compared with the original cutoffs based on prevalence rates for overweight and obesity from US and Chinese data.

Table 10.12. SD score cutoffs corresponding to the international BMI cutoffs. Note the new cutoff for morbid obesity (BMI=35). From Cole and Lobstein (2012).

BMI cutoff at 18y (kg/m^2)	SD score equivalent	Percentile equivalent
Boys		
25	1.310	90.5
30	2.288	98.9
35	2.930	99.83
Girls		
25	1.244	89.3
30	2.192	96.6
35	2.822	99.76

A worked example on how to derive these new cutoffs for obesity in children is given in the appendix of Cole and Lobstein (2012), allowing researchers to construct their own cutoffs for any required BMI (e.g., BMI 23 for Asians) at 18y. The adoption of this reformulated approach has several important advantages. In summary:

- Use of weight-for-stature charts, that cover only a restricted age range, and omit adolescence entirely, are eliminated
- Obesity assessment in adults is integrated with that in children and adolescents
- Existing cutoffs can now be expressed as percentiles or SD scores allowing comparison with other BMI references
- New cutoffs are easy to derive (e.g., BMI for morbid obesity)

10.4.2 WHO classification of over-weight and obesity in children

With the development of international growth references for both young children 0–5y and school-aged children and adolescents for clinical and epidemiological use, WHO has adopted cutoffs points for overweight and obesity for the two age groups based on standard deviation scores (i.e., Z-scores) derived from the corresponding BMI-for-age curves of the corresponding reference data.

For infants and children 0–5y, the WHO Child Growth Standards based on the multicenter study of children aged 0–5y from 6 diverse geographic sites (Brazil, Ghana, India, Norway, Oman, and the United States) should be used (WHO, 2006). whereas, for school-age children and adolescents the WHO international growth reference 2007 is recommended. The latter is a reconstruction of the original 1977 National Center for Health Statistics (NCHS) data set supplemented with data from the WHO Child Growth Standard to ensure a smooth transient at age 5y (de Onis et al., 2007). The statistical methodology used to construct this reference was the same as that used for the WHO Child Growth Standard (2004). The full set of tables and charts for the BMI-for-age curves by sex for both preschool and school-aged children, including application tools, is available (WHO international growth reference 2007).

For a child 0–5y, WHO defines "at risk of overweight" as a BMI Z-score of > +1 but < +2; "overweight" as a BMI Z-score > +2 and < +3, and "obese" as a BMI Z-score > +3, based on the WHO Child Growth Standard (WHO, 2006; de Onis & Lobstein, 2010). For a school-aged child aged 5–19y, WHO defines a BMI Z-score above +1 but < +2 as overweight, and +2 as obesity, based on the WHO 2007 growth reference data (de Onis et al., 2007;de Onis & Lobstein, 2010).

Details for the justification of these WHO cutoffs were not provided. The more stringent WHO cutoffs for children aged < 5y compared to those ≥ 5y were proposed to avoid overdiagnosis of obesity in younger children; the increase in average BMI observed in all age groups of children has been greater in the older children since the 1990s. Hence, the current trend in BMI in children no longer reflects what is termed "age agnosticism", in which the risk of obesity is assumed to be constant throughout infancy and childhood (de Onis & Lobstein, 2010; Wright et al., 2022).

Cole and Lobstein (2012) compared the prevalence rates for overweight and obesity defined by the WHO BMI Z-scores and their extended international IOFT cutoffs using US and Chinese data. There were systematic differences in the prevalence of overweight and obesity; the WHO prevalence rates were lower for children age 2–5y, but higher for those 5–18y, probably due to the method of construction of the WHO percentile curves; see Cole and Lobstein (2012) for more details.

10.4.3 U.S. classification of overweight and obesity in children

The approach of Cole et al. (2000; 2012) has not been adopted in the United States. Instead, the U.S 2000 CDC growth charts are used, with overweight among children aged 2–19y defined as BMI > 85th but < 95th percentile, and obesity as BMI > 95th percentile of the sex-specific CDC BMI-for-age growth charts. Two sets of clinical charts for children are available. Set 1 has the outer limits of the curves at the 5th and 95th smoothed percentiles and includes the 85th percentile for BMI-for-age as shown in Figure 10.27. Set 2 has the outer limits of the curves at the 3rd and 97th percentiles and are for specialist use. For details on the compilation of the U.S 2000 CDC growth charts, see Section 10.7. For details of the BMI-for-age charts see: 2000 CDC growth charts.

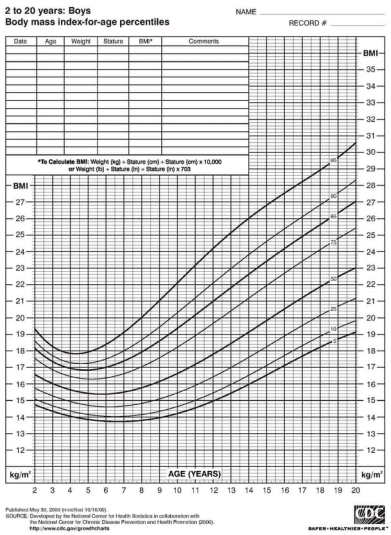


Figure 10.27.U.S 2000 CDC Body mass index-for-age percentiles: boys 2–20y.

The smoothed percentiles of the original U.S CDC reference growth charts extended only up to the 97th percentile. The marked rise in the prevalence of extreme obesity among children in the U.S (i.e, BMI > 99th percentile of the CDC growth charts) has led to alternative definitions of severe obesity in U.S children aged > 2–19y. They include: BMI > 120% of the 95th percentile (1.2 × 95th percentile) for age, or an absolute BMI > 35kg/m^2 for adolescents, whichever is lower based on age and sex. The inclusion of an absolute BMI > 35kg/m^2 threshold was chosen by the American Heart Association because it aligns the pediatric definition of severe obesity with the U.S class II obesity in adults (Table 10.7), in whom there is a high risk of early mortality (Kelly et al., 2013). However, these alternative definitions of severe obesity in children aged > 2 to 19y have some limitations which have been discussed in detail in a recent CDC report by Hales et al. (2022). CDC has now extended their BMI distribution on their BMI-for-age growth charts by using additional data from NHANES (1999–2016), rather than relying on extrapolation. These BMI-for-age charts extend up to a BMI of 60kg/m^2 with four additional percentile curves above the 95th (i.e., 98th, 99th, 99.9th, and 99.99th) by gender.

The CDC (2022) recommend using these extended BMI-for-age growth charts (Figure 10.28) for monitoring obese children and adolescents in place of the severe obesity growth charts that are used in clinical care (Gulati et al., 2012; Kelly et al., 2013). For monitoring growth of U.S children without obesity, however, the 2000 CDC BMI-for-age growth charts (e.g., Figure 10.27) should be used.

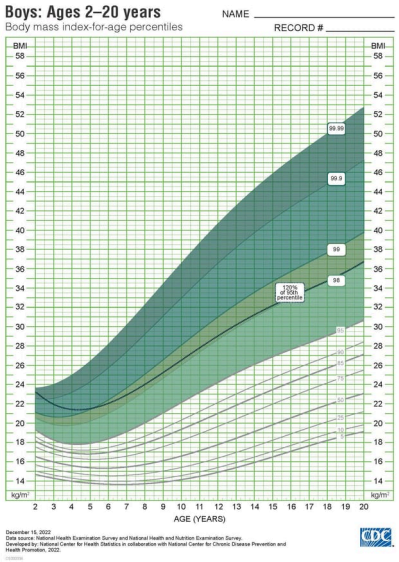


Figure 10.28. Extended CDC BMI-for-age growth chart for boys. From CDC (2022)

Unlike the two different cutoffs for children aged < 5y versus ≥ 5y set by WHO (de Onis et al., 2007), the CDC percentile cutoffs assume that the risk of obesity is constant throughout childhood, thus reflecting “age agnosticism”, an approach suggested to result in overdiagnosis of obesity in younger children (Wright et al., 2022). However, the CDC does caution that “in-depth assessments” are required to determine whether children and adolescents with BMI-for-age > 95th percentile are truly overfat, and at increased risk for health complications.

Skinfold thickness and waist circumference have been proposed as alternatives for quantifying adiposity in overweight and obese children and adolescents but are challenging in children with severe obesity and not recommended. For a review of the existing evidence for an association of excess adiposity with cardiovascular and metabolic risk in children, see Chung et al. (2018).

10.4.4 Comparisons of overweight and obesity in children among countries

BMI is accepted as a valid indirect measure of adiposity in children. However, unlike adults, the 50th percentile BMI varies with age as well as weight. Consequently, BMI values in childhood must be compared with age and sex-specific reference data sets. Unfortunately, differences exist in the methods that have been used to construct the curves for the reference data sets, some of which are based on representative national data (e.g., 2000 CDC reference data), whereas others provide international reference data (Cole et al., 2000; WHO, 2006). Marked variations across countries also occur in the terminology used to define the levels of BMI indicative of overweight and obesity. For example, based on the French BMI reference data, “overweight” corresponds to values above the 97th percentile and no cutoff for “obesity” exists (Rolland-Cachera et al., 2002), whereas the US CDC now define overweight as a BMI ≥ 85th percentile but < 95th percentile, and obesity as ≥ 95th percentile. In contrast, WHO apply Z-scores to indicate “risk of overweight”, “overweight”, and “obesity”, and these differ for children less than and greater than aged 5y (Section 10.4.2).

As a result of these differences, comparison of the prevalence of overweight and obesity across countries is challenging, and the prevalence estimates differ even when compared in the same population of children. For example, estimates often appear highest when using the WHO classification systems, intermediate with the 2000 CDC definition, and lowest based on the IOTF definition (Shields & Tremblay, 2010).

An additional challenge complicating the interpretation of the data is the absence of international validated gold-standard upper BMI cutoffs for overweight and obesity in childhood. The existing prospective studies in which excess adiposity in children has been associated with risk of cardiovascular disease are limited by small sample sizes (Chung et al., 2018). Moreover, to date, there are no large-scale longitudinal data that relate overweight in childhood with adverse health outcomes in adulthood because of the long-time span required before they appear (i.e., middle-age or beyond) (Wright et al., 2022). This has led to the conventional

statistical definitions for the BMI cutoffs for overweight and obesity, with the assumption that the prevalence of overweight or obesity is the same for each age- and sex-specific group, even though it appears that older children now experience a greater increase in BMI than younger children.

Clearly, an international consensus is necessary before meaningful comparisons between studies of overweight and obesity during childhood can be made (Rolland-Cachera, 2011; Flegel & Ogden, 2011). The current recommendation by the European Childhood Obesity Group (ECOG) is to use the IOTF and WHO definitions for international comparisons of the prevalence of childhood overweight and obesity, despite their limitations. For comparison within countries, however, national age- and sex-specific BMI reference data and definitions of overweight and obesity are appropriate(Rolland-Cachera, 2011). Further discussion on the use of international versus national BMI reference data are available in Chinn & Rona (2002) and Reilly et al. (2010).

Unfortunately, the evidence to identify children who are truly at risk for future adverse health outcomes arising from overweight or obesity remains insufficient. It appears that the conventional statistical approach to define overweight and obesity is no longer valid and results in overdiagnosis of obesity in younger children when a true measure of total fat mass is determined based on dual energy x-ray absorptiometry (Vanderwall et al., 2017)or deuterium dilution (Wright et al., 2022). Therefore, more research is urgently needed to improve the diagnosis of childhood obesity and define its adverse health consequences.

10.4.5 Using BMI to define thinness in children and adolescents

The term “thinness” was adopted by Cole et al. (2007) to describe a low BMI to avoid confusion with the terms “wasting” (defined by a weight-for-stature< –2 Z-score) and “underweight” (defined by a weight-for- age < –2 Z-score).

Cole et al. (2007) used the same methods to define grades of thinness in children and adolescents as they used previously to define overweight and obesity in this age group (Section 10.4.1). For each dataset, percentile curves were drawn to pass through the BMI cutoff values of 18.5, 17, and 16kg/m^2 at age 18y, consistent with the grades 1, 2, and 3 of thinness used by WHO for adults (Table 10.12). The percentile curve passing through a BMI value of 17kg/m^2 at age 18y gave a mean BMI close to a Z-score of –2, matching the existing WHO criteria for wasting in children less than 5y of age (i.e., weight-for-height below –2 Z-score). Hence, the authors proposed that this percentile curve should be a basis for an international definition of thinness in children and adolescents.

However, a major disadvantage of these childhood BMI cutoffs for thinness is that they are not expressible as BMI percentiles. Consequently, Cole and Lobstein reformulated their international cutoffs for thinness for children aged 2–18y so that BMI could be expressed as percentiles or Z-scores for comparison with the WHO BMI cutoffs (Cole & Lobstein, 2022). Table 10.13 shows the sex-specific Z-score cutoffs and their corresponding percentile equivalents to the original international BMI cutoffs (16, 17, and 18.5kg/m^2 at age 18y) which are consistent with the grades 3, 2, and 1 of thinness used by WHO for adults (Section 10.3.5). For a comparison of these new cutoffs for thinness with the originals, see Cole & Lobstein (2012).

Table 10.13. SD score cutoffs corresponding to the international BMI cutoffs for thinness. From Cole and Lobstein (2012).

BMI cutoff at 18y (kg/m^2)	SD score equivalent	Percentile equivalent
Boys		
16	-2.565	0.52
17	-1877	3.0
18.5	-1.014	15.5
Girls		
16	-2.436	0.74
17	-1.789	3.7

BMI cutoff at 18y (kg/m^2)	SD score equivalent	Percentile equivalent
18.5	-0.975	16.5

The BMI-for-age reference data as sex-specific Z-scores or as percentiles for both children aged 0–5y and school-aged children and adolescents (5–19y) are available as charts and tables from WHO (Child growth standards; WHO international growth reference 2007). Thinness and severe thinness are defined as: BMI –2 Z-score and BMI –3 Z-scores, respectively. Raw measurements can be calculated as BMI-for-age percentiles and Z-scores using the WHO AnthroPlus application (AnthroPlus 2009).

In an effort to avoid overestimating the prevalence of thinness among South Asian children, de Wilde et al. (2013) have suggested that ethnic specific BMI-cutoffs for thinness should be developed, as proposed for the BMI cutoffs for overweight and obesity.

This page titled 9.5: BMI in children and adolescents (10.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

10: BODY COMPOSITION (CHAPTER 11)

Abstract

Most anthropometric methods used to assess body composition are based on the two componment model whereby the body consists of fat and fat-free mass. These two body components can be assessed indirectly from selected skinfold thickness and circumference measurements taken by standardized techniques. Several methods exist for estimating percentage body fat and/or total body fat. In the simplest method, skinfold thickness measurements, either singly or in combination, are used to assess body fat (as % or total). Alternatively, percentage body fat can be predicted from adiposity equations matched to the measured skinfolds and study population (by age, gender, ethnicity, activity level, etc). Arm-fat area, calculated from triceps skinfold thickness and mid-upper arm circumference (MUAC), is also used as a proxy for total body fat, although the equation used has some limitations. Both WHO international and population-specific reference data are available for triceps and subscapular skinfolds for children. Arm-fat area data are more limited, although data for U.S. children (1-20y) from the CDC2000 BMI growth chart sample have been compiled.

Anthropometric variables from multiple anatomical sites are also used to estimate body density, from which the percentage of body fat, and subsequently, total body fat is calculated. The reliability of this method has been questioned based on comparisons of the derived percentage body fat estimates against those generated using the in vivo gold standard 4-component model which does not rely on any theoretical assumptions. Corrections that account for age, sex, disease state or nutritional status can now be applied to the density-based formulae and/or the empirical equations used to relate fat content to body density, and thus improve the final assessment of body composition.

Recognition of the link between the distribution of body fat and risk of cardiovascular disease has prompted use of waist-hip ratio (WHR), and more recently, waist circumference (WC), as practical anthropometric surrogates for intra-abdominal visceral fat. Population-specific cutoffs for adults have been set to denote high WHR or WC indicative of abdominal obesity and cardiovascular risk. Increasingly, WC is being included along with BMI in all obesity surveillance studies.

Fat-free mass can be estimated as body weight (kg) minus body fat from the adiposity or density-based methods outlined above. Alternatively, simpler methods include the measurement of MUAC, either alone or combined with triceps skinfold thickness to calculate arm mid-upper-muscle circumference (MUAMC) or arm muscle area (AMA). MUAC alone is used in emergencies to screen for severe acute malnutrition (SAM), whereas MUAMC and AMA can be used as proxies for muscle mass, and thus for the detection of sarcopenia in the elderly. AMA is preferable to MUAMC because it more adequately reflects the true magnitude of tissue changes.

Calf circumference and hand grip strength as surrogate markers of skeletal muscle mass and strength respectively, are increasingly being used, in part because loss of both muscle mass and strength has been associated with several adverse health outcomes. Both measurements are recommended by the Asian and European Working Groups on Sarcopenia to identify at risk older adults. The measurements are also used in children and athletes to assess physical fitness. Adiposity has been identified as a confounder of calf circumference measurements, and BMI adjustment factors have been developed. Low calf circumference with any BMI can now be identified. Population-specific calf circumference cutoff values are available to detect low muscle mass in adults. Handgrip strength is said to be a better predictor of functional health outcomes than low muscle mass, and has been linked with alteration in physical performance in cross-sectional studies. Associations with all-cause mortality, cardiovascular mortality, and hospital readmissions have also been observed in prospective cohort studies. Hand grip strength is measured with a calibrated handheld dynamometer. The measurements depend on the model used, but dynamometer-specific cutoffs values are not yet applied. Current cutoffs for weak muscle strength vary across regions; population specific normative reference data are available for the elderly and across the life course.

Finally, the cross-sectional nature of the normative reference data compiled for the anthropometric variables discussed limits their use for monitoring the trajectories of individuals and the degree to which causal and age-related inferences can be drawn. None of the anthropometric variables are sensitive enough to monitor small changes in body fat or fat-free mass that may arise after short-term nutritional support or deprivation.

10.1: Anthropometric assessment of body composition (11.0)

10.2: Assessment of body fat (11.1)
10.3: Assessment of fat-free mass (11.2)

This page titled 10: Body composition (Chapter 11) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Anthropometric assessment of body composition (11.0)

Most anthropometric methods used to assess body composition are based on a model in which the body consists of two chemically distinct components: fat and the fat-free mass, with the principle that if one of these components is determined, the other can be estimated. The amount and distribution of both body fat and the fat-free mass have important health outcomes in infants, children, and adults.

Fat is the main long-term storage form of energy in the body, and alterations in body fat content provide indirect estimates of changes in energy balance. Most of the body fat is stored in adipose tissue, which is distributed in different proportions throughout the body. The pattern of distribution of adipose tissue is dependent on many factors including sex, age, race/ethnicity, genotype, diet, physical activity, and hormone levels. Adipose tissue is traditionally distributed into two main components, each with different metabolic characteristics: subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT). Of the two components, visceral adipose tissue is a hormonally active component of total body fat tissue and an independent risk marker of cardiovascular and metabolic morbidity and mortality. Abnormally high deposition of VAT is known as visceral obesity (Neeland et al., 2019).

Fat may also be present in areas of the body where fat is not physiologically stored, such as liver, pancreas, heart, and skeletal muscle. Fat surrounding these organs is termed ectopic fat, and its deposition might contribute to increased risk of atherosclerosis and type 2 diabetes (Neeland et al., 2019). The mechanisms whereby an excess of VAT is related to various health outcomes, as well as the tendency to deposit adipose tissue in ectopic depots are not fully understood; see Neeland et al., 2019 for more details.

The fat-free mass consists of the skeletal muscle, non-skeletal muscle, organs, connective tissue, total body water, and the skeleton. Muscle is a major component of the fat-free mass and the primary site for glucose uptake and storage, as well as a reservoir of amino acids stored as protein. Loss of muscle mass is associated with several negative health outcomes, including delayed recovery from illness, slowed wound healing, reduced metabolic rate, and physical disability (Argilés et al., 2016).

Patients at high risk of losing muscle are those who are experiencing weight loss through diseases or conditions associated with inflammatory components, and malnutrition (Cruz-Jentoft et al., 2019). Risk of death during infections is exacerbated by the loss of muscle mass in malnourished children (Briend et al., 2015). Aging may also lead to a loss of muscle mass. This condition is known as sarcopenia, and results in diminished quality of life, greater susceptibility to infection, and an increased risk of mortality (Deutz et al., 2019).

The anthropometric measurements of body composition are fast, noninvasive, and require the minimum of equipment compared to laboratory techniques. Consequently, anthropometry has been the method most frequently used in the past in both routine clinical and public health settings. Increasingly, however, recognition of the limitations of anthropometry to assess body composition has led to the use of alternative laboratory methods to assess body composition such as bioelectrical impedance analysis (BIA) and dual energy X-ray absorptiometry (DXA) in these settings (Howell et al., 2018; Schubert et al., 2019). See Chapter 14 for details of these laboratory methods.

In clinical practice, indices of body composition can be used to identify patients with chronic under- or overnutrition and to monitor long-term changes in body composition during nutritional support. In public health, they can identify individuals who are vulnerable to under- or overnutrition and help evaluate the effectiveness of nutrition intervention programs.

Details of the standardized procedures used for anthropometric measurements of body composition, and the derivation of the more important indices and their limitations are summarized in this chapter and are given in detail in Lohman et al. (1988). Variability in body composition across populations associated with lifestyle, environment, genetics, and ethnicity has emphasized the need for population-specific reference data to interpret body composition indices (Wells, 2019). Hence, where appropriate, the interpretive criteria available for the assessment of body composition based on anthropometry, are also summarized. For a review of the methods used to develop reference values and cutoff points, the reader is advised to consult Chapter 1.

This page titled 10.1: Anthropometric assessment of body composition (11.0) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

10.2: Assessment of body fat (11.1)

The body fat content is the most variable component of the body, differing among individuals of the same sex, height, and weight. Estimates of total body fat, together with the rate of change in the body fat content, are often used to assess the presence and severity of undernutrition. A large and rapid loss of body fat is indicative of severe negative energy balance. Small changes in body fat (i.e., < 0.5kg), however, cannot be measured accurately using anthropometry.

On average, the fat content of women is higher than that of men, representing 26.9% of their total body weight compared with 14.7% for men (Table 11.1). Body fat is deposited in two major types of sites: one for essential lipids, and the other for storage of fat. Essential lipids are found in the bone marrow, central nervous system, mammary glands, and other organs and are required for normal physiological functioning; fat from these sites makes up about 9% (4.9kg) of body weight in reference woman and 3% (2.1kg) in reference man. Storage fat consists of inter- and intra-muscular fat, fat surrounding the organs (e.g., liver, heart, pancreas) and gastrointestinal tract, and subcutaneous fat (Lohman, 1981). The proportion of storage fat in males and females is relatively constant, averaging 12% of total body weight in males and 15% in females.

Of the total body fat, over one-third in reference man and woman is estimated to be subcutaneous fat. Body fat is expressed either in absolute terms (the weight of total body fat in kilograms) or as a percentage of the total body weight. There is, however, a lack of consensus about the usefulness of percentage body fat as an index of adiposity. Some investigators argue that percentage body fat over-adjusts for weight because it includes the fat mass component in both the numerator and denominator (Cole et al., 2008). A further limitation is that percentage fat is not fully independent of body size. High percentage fat values might reflect high adiposity or low lean mass (Wells, 2014).

Table 11.1 Distribution of body fat in reference man and women. Data in kilograms. Weights for total fat and body weight in reference et al.(1956), Alexander (1964), and Wilmore and Brown (1974).

Fat location	man
Essential fat (lipids of the bone marrow, central nervous systems, mammary glands and other organs)	2.1
Storage fat (depot)	8.2
Subcutaneous	3.1
Intermuscular	3.3
Intramuscular	0.8
Fat of thoracic and abdominal cavity	1.0
Total fat	10.5
Body weight	70.0
Percentage fat	14.7

In population studies, body fat is often assessed by anthropometry. In the past, body mass index has been the principal index used to predict excess adiposity; see Chapter 10 for more details. However, skinfold thickness determinations, either alone or in association with other anthropometric variables (e.g., limb girths and breadths), are also used to predict percentage body fat; over six hundred prediction equations have been developed. Associations between these anthropometric variables and percent body fat differ by many factors including gender, age, race / ethnicity, and level of adiposity so that prediction equations must be carefully matched with the population under study and standardized techniques used for the measurements (Provyn et al., 2011).

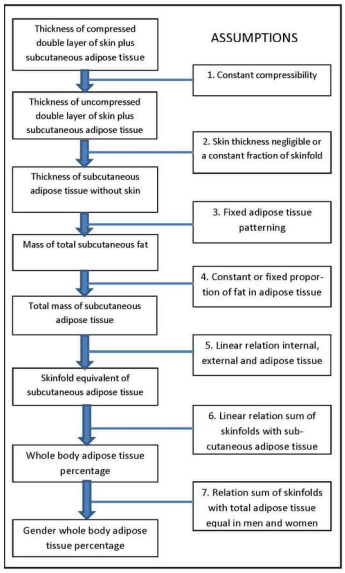


Figure 11.1. Flow chart of the transformation from skinfold to total body adiposity; eight possible steps (left) and possible assumptions (right). Modified from: Provyn et al. (2011)

More recently, the importance of the distribution of body fat has been emphasized. Numerous studies have reported correlations between the amount of intra-abdominal fat (i.e., visceral adipose tissue) and metabolic disturbances linked to the risk of cardiovascular disease (Neeland et al., 2019; Ross et al., 2020). These findings have led to the assessment of visceral adipose tissue as an independent risk marker of cardiovascular and metabolic morbidity and mortality (Hiuge-Shimizu et al., 2012). Waist-hip circumference ratio and, increasingly, waist circumference alone, are being used as anthropometric surrogates for intra-abdominal visceral fat (Sections 11.1.6 and 11.1.7).

11.1.1 Skinfold thickness measurements

Skinfold thickness measurements provide an estimate of the size of the subcutaneous fat depot, which, in turn, has been used to derive an estimate of total body adiposity. Such an estimate is based on seven assumptions shown in Figure 11.1, most of which are not true. For example, the relationship between subcutaneous and internal fat is nonlinear and varies with body weight and age: very lean subjects have a smaller proportion of body fat deposited subcutaneously than obese subjects. Moreover, variations in the distribution of subcutaneous fat occur with sex, race or ethnicity, and age (Wagner & Heyward, 2000). For a detailed discussion of the limitations of each of the seven assumptions depicted in Figure 11.1, see Provyn et al. (2011).

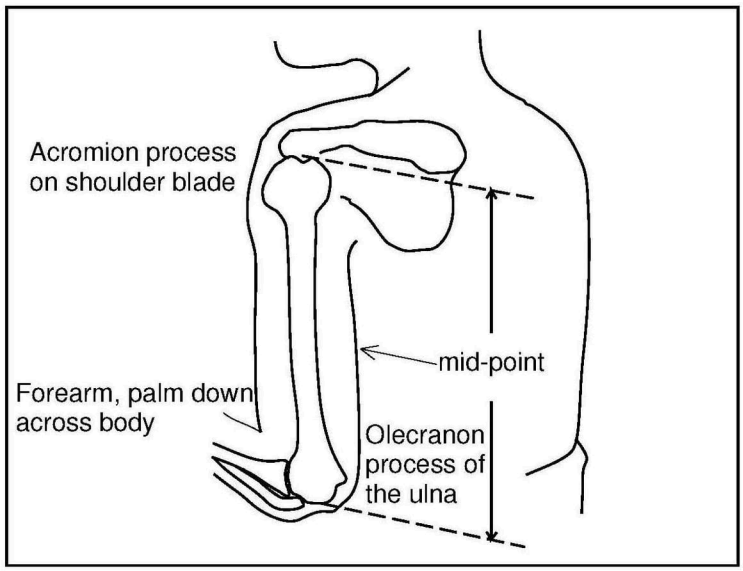


Figure 11.2. Location of the midpoint of the upper arm. Redrawn from Robbins et al.(1984).

The following skinfold sites, described in detail in Lohman et al. (1988),are commonly used:

- Triceps skinfold is measured at the midpoint of the back of the upper arm (Figure 11.2).
- Biceps skinfold is measured as the thickness of a vertical fold on the front of the upper arm, directly above the center of the cubital fossa, at the same level as the triceps skinfold.
- Subscapular skinfold is measured below and laterally to the angle of the shoulder blade, with the shoulder and arm relaxed. Placing the subject's arm behind the back may assist in identification of the site. The skinfold should angle 45° from horizontal, in the same direction as the inner border of the scapula (i.e., medially upward and laterally downward) (Figure 11.3A).

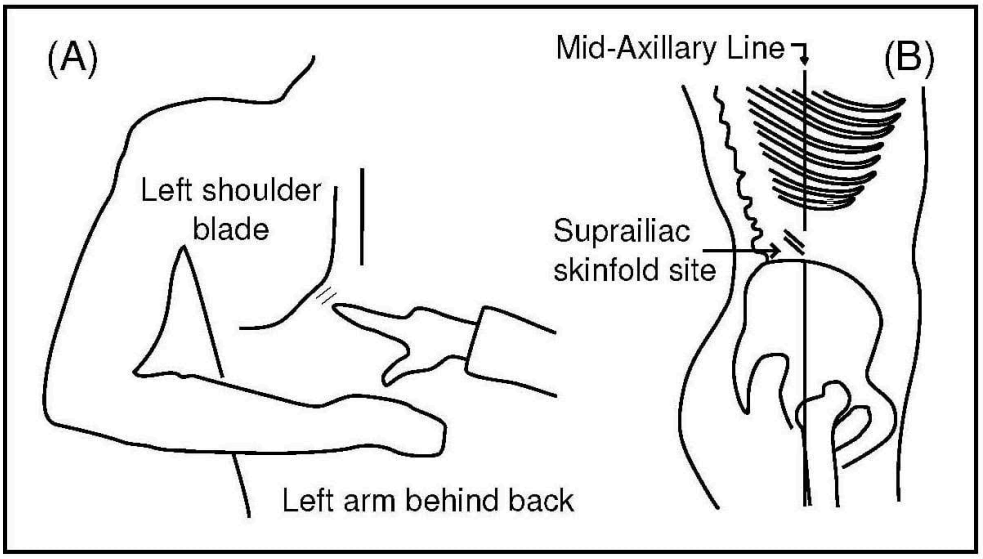


Figure 11.3. Location of the subscapular (A) and suprailiac (B) skinfold sites.

- Suprailiac skinfold is measured in the midaxillary line immediately superior to the iliac crest. The skinfold is picked up obliquely just posterior to the midaxillary line and parallel to the cleavage lines of the skin (Figure 11.3B).
- Midaxillary skinfold is picked up horizontally on the midaxillary line, at the level of the xiphoid process.

Marked ethnic differences in adiposity based on skinfolds (as well as fat mass via BIA and DXA) have been reported. For example, greater subcutaneous fatness was reported for white boys compared to their black counterparts in the U.S. (Addo & Himes, 2010). These data were based on the population of healthy U.S. children aged 1-20y used to construct the CDC 2000 BMI charts (Kuczmarski et al., 2000a). For these BMI charts, the weight data for children > 6y who participated in the NHANES III survey were excluded because the inclusion of these data shifted the upper percentile curves.

Race-ethnicity differences in skinfolds have also been reported in children living in the U.K. Adiposity levels were higher among South Asian children based on the sum of four skinfolds (biceps, triceps, subscapular and suprailiac), whereas black African Caribbean children had similar or lower adiposity levels than white Europeans (Nightingale et al., 2011). Clearly, race-ethnicity differences in fat patterning should be taken into account when interpreting results based on subcutaneous skinfolds.

Skinfold thickness measurements are best made using precision thickness calipers; they measure the compressed double fold of fat plus skin. As a result of the compression, they always underestimate actual subcutaneous fat thickness. The skinfold is always grasped at the marked site with the fingers on top, thumb below, and forefinger on the marked site. Three types of precision calipers can be used: Harpenden, Lange, and Holtain (Figure 11.4).

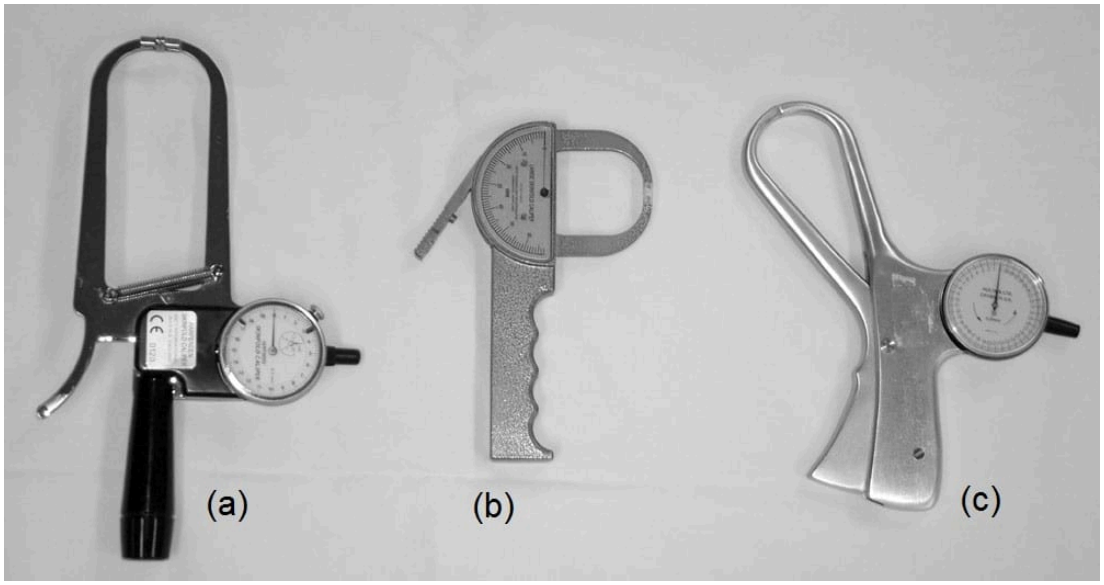


Figure 11.4. Harpenden (a), Lange (b), and Holtain (c) precision skinfold thickness calipers.

Precision calipers are designed to exert a defined and constant pressure throughout the range of measured skinfolds and to have a standard contact surface area or “pinch” area of 20-40mm^2. The skinfold calipers must be recalibrated at regular intervals using a calibration block. Both the Harpenden and Holtain skinfold calipers, which have a standard jaw pressure of 10g/mm^2, give smaller skinfold values than Lange calipers, which are fitted with a lighter spring (Gruber et al., 1990). For example, values from Holtain calipers are about 2-5mm (mean) lower than those obtained using the Lange calipers (Lohman et al., 1984). Hence, care must be taken to ensure the same precision calipers are used when examining secular trends in skinfold thicknesses.

For all the skinfold measurements, the subject should stand erect with the weight evenly distributed and feet together, shoulders relaxed, and arms hanging freely at the sides. The measurement technique is described in detail for the triceps skinfold, as the latter is the site most frequently used to obtain a single indirect measure of body fat; the technique used for the other skinfold sites is similar. There is no consensus as to whether the left or right side of the body should be used. In the WHO Multicentre Child Growth Reference Study, triceps and subscapular measurements were taken on the left side of the body (de Onis et al., 2004). A description of these measurement protocols are available in the WHO anthropometric training video. However, the current practice of the U.S. National Health and Nutrition Examination Surveys (NHANES) is that skinfold sites are measured on the right side of the body.

Measurement of triceps skinfold

The measurement of the triceps skinfold is performed at the midpoint of the upper right arm, between the acromion process and the tip of the olecranon, with the arm hanging relaxed. To mark the midpoint, the right arm is bent 90° at the elbow, and the forearm is placed palm down across the body. Then the tip of the acromion process of the shoulder blade at the outermost edge of the shoulder and the tip of the olecranon process of the ulna are located and marked. The distance between these two points is measured using a non-stretchable tape, and the midpoint is marked with a soft pen or indelible pencil, directly in line with the point of the elbow and acromion process (Figure 11.2). The right arm is then extended so that it is hanging loosely by the side. The examiner grasps a vertical fold of skin plus the underlying fat, 2cm above the marked midpoint, in line with the tip of the olecranon process, using both the thumb and forefinger. The skinfold is gently pulled away from the underlying muscle tissue, and then the caliper jaws are applied at right angles, exactly at the marked midpoint (Figure 11.5). The skinfold remains held between the fingers while the measurement is taken.

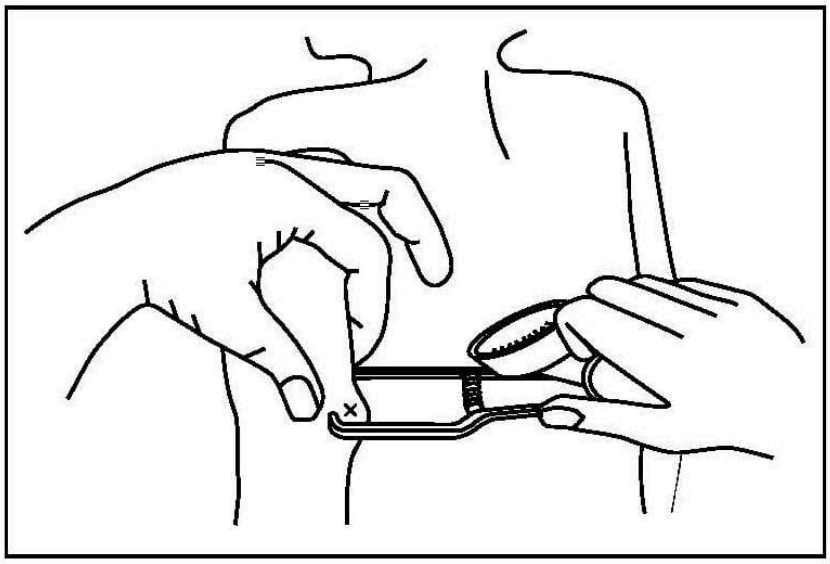


Figure 11.5 Measurement of the triceps skinfold in the upright position using the Harpenden caliper. Redrawn from: Robbins et al. (1984).

When using the Lange, Harpenden, or Holtain calipers, pressure must be applied to open the jaws before the instrument is placed on the skinfold; the jaws will then close under spring pressure. As the jaws compress the tissue, the caliper reading generally diminishes for 2-3s, and then the measurements are taken. Skinfolds should be recorded to 0.1mm on the Harpenden and Holtain skinfold calipers and to 0.5mm on the Lange.

Triceps skinfold measurements can also be made with the subject lying down. The subject lies on the left side with legs bent, the head supported by a pillow, and the left hand tucked under the pillow. The right arm rests along the trunk, with the palm down. The measurement is taken at the marked midpoint of the back of the upper right arm, as described above. The examiner should be careful to avoid parallax errors by bending down to read the calipers while taking the measurements (Chumlea et al.,1984).

Precision of skinfold measurements

Within-examiner and between-examiner measurement errors can occur when measuring skinfolds, particularly for subjects with flabby, easily compressible tissue or with very firm tissue that is not easily deformed(Lukaski, 1987). Errors may also occur when measuring skinfolds in obese subjects (Forbes et al., 1988).

Within-examiner errors can occur when the same examiner fails to obtain identical results on repeated skinfolds on the same subject; such errors are a function of the skinfold site, the experience of the examiner, and the fatness of the subject. Within-examiner measurement errors can be small when measuring triceps skinfolds, provided that training in standardized procedures

is given; the errors in these circumstances typically range from 0.70-0.95mm (Table 11.2).

Table 11.2 Reported values for within-observer and between-observer technical error of the measurement (TEM) for skinfold measure Ulijaszek & Kerr (1999).

Skinfold measurement	no. of studies	Mean (mm)	Range
Within-observer TEM			
Biceps	3	0.17	0.1–0.2
Triceps	21	0.84	0.1–3.7
Subscapular	19	1.26	0.1–7.4
Suprailiac	10	1.16	0.1–3.2
Between-observer TEM			
Biceps	8	0.84	0.2–2.1
Triceps	28	1.06	0.2–4.7
Subscapular	28	1.21	0.1–3.3
Suprailiac	11	2.28	0.3–6.4

Between-examiner errors arise when two or more examiners measure the same subject and skinfold site; such errors are usually larger than within-examiner errors, but they can be reduced to not more than 2mm with training and care (Burkinshaw et al., 1973). Within- and between-examiner measurement errors tend to be greater if very large (> 15mm) or small (< 5mm) skinfolds are measured (Edwards et al., 1955).

Table 11.2 lists some reported values for both within- and between-examiner technical error of the measurement (TEM) (Chapter 9) for biceps, triceps, subscapular, and suprailiac skinfold measurements, compiled by Ulijaszek and Kerr (1999). Consult Chapter 9 on how to measure TEM.

Within- and between examiner TEMs for triceps and subscapular skinfolds were also calculated in the WHO Multicentre Growth Reference Study (MGRS) (de Onis et al., 2004); the values are shown in (Table 11.3). As expected, the range for the between-examiner TEM for both the longitudinal and cross-sectional components of the MGRS from the six country sites was larger than the range for the within-examiner TEM for these two skinfolds. For more details, see WHO (2006).

Zerfas (1985) has evaluated the measurement error for skinfolds from any site using a repeat-measures protocol and recommended target values for the differences between the trainee and a criterion

Table 11.3 Reported values in mm for the within-examiner and between-examiner technical error of the measurement (TEM) for the routine MGRS data. Longitudinal measurements were made by the follow-up team during the longitudinal component. Cross-sectional data are from the MGRS cross-sectional component.

Skinfold measurement	Range (mm)
Within-examiner TEM	MGRS teams
Triceps	0.39-0.61
Subscapular	0.29-0.41
Between-examiner TEM	
Triceps	
Longitudinal	0.50-0.83

Skinfold measurement	Range (mm)
Cross-sectional	0.46-0.85
Subscapular	
Longitudinal	0.42-0.69
Cross-sectional	0.44-0.62

anthropometrist; the target training values are shown in (Table 11.4). A difference of more than 5mm between the measurements of the criterion anthropometrist and the trainee indicates a gross error related to the reading or recording; a difference between the measurement of the criterion anthropometrist and the trainee of 0.0-0.9mm indicates that the trainee has reached an acceptable level of proficiency in the measurement technique.

In the WHO Multicentre Growth Reference Study, measurements for triceps and subscapular skinfolds were taken on each child by two trained and standardized anthropometrists. Their values were then compared to ensure that the duplicate measurements were within the maximum allowable difference, designated as 2.0mm for each skinfold(de Onis et al., 2004).

Sports anthropometrists have set target values for training which also include skinfolds and arm circumference measurements (Gore et al., 1996); these could be adopted by nutritionists. Suggested target values are expressed as TEM (as a percentage), and for skinfolds are 7.5 (level 1) and 5.0 (levels 2 and 3). Criterion anthropometrists should be expected to achieve a %TEM of 5.0 for skinfolds.

Table 11.4 Evaluation of measurement error in anthropometric measurements. After Zerfas (1985). Differences greater than those nc under "Poor" are taken to indicate a gross error. Data from Ulijaszek & Kerr (1999).

	Trainee-trainer difference		
Measurement	Good	Fair	Poor
Height or length (mm)	0–5	6–9	10–19
Weight (kg)	0–0.1	0.2	0.3–0.4
Arm circ. (mm)	0–5	6–9	10–19
skinfolds (any) (mm)	0–0.9	1.0–1.9	2.0–4.9

Secular trends in adiposity across populations have been examined by measuring triceps and subscapular skinfold thicknesses. However, in a sample of > 45,000 U.S. adults participating in the NHANES surveys conducted from 1988-1994 through 2009-2010, Freedman et al. (2017) concluded that it is unlikely that skinfold thicknesses could be used to monitor trends in obesity. The changes in the measured skinfold thicknesses were small and fell within the technical error of the respective skinfold measurements.

Interpretive criteria for triceps and sub-scapular skinfolds

The WHO included triceps and subscapular skinfold thickness measurements in the construction of the Multicenter Child Growth Standard (MCGS) for young children aged 0-5y. Children from six diverse countries (Brazil, China, India, Norway, Oman, and the USA) were included. To reduce the impact of environmental variation, only privileged healthy populations were selected (See Chapters 9 and 10 for more details). Charts based on sex-specific percentiles and Z-scores for triceps-for-age (WHO MCGS Triceps) and subscapular-for-age (WHO MCGS Subscapular) are available for children 3mos-5y. Details of the standardized methods used and the development of these reference data are available (de Onis et al., 2004).

Age- and sex-standardized percentile reference curves for triceps and subscapular skinfold thicknesses have also been compiled for children of varying ages in several high-income countries (e.g., U.S., Spain, Poland) (Addo & Himes, 2010; Moreno et al., 2007; Jaworski et al., 2012). In the United States numerical data for the smoothed percentiles for triceps and subscapular skin-

folds for U.S. girls and boys aged 1.50-19.99y are available in Addo and Himes (2010). These reference data are based on the same population of children and adolescents used to construct the CDC 2000 growth curves for BMI-for-age (Kuczmarski et al., 2000a). Serrano et al. (2015) have cautioned the use of these U.S. skinfold percentiles for interpreting skinfolds from Hispanic American children and adolescents because schoolchildren from Spain, Argentina, Cuba, Venezuela and Mexico were found to have higher triceps and subscapular percentiles than those of the CDC reference (Addo & Himes, 2010; Kuczmarski et al., 2000a). Instead, Serrano et al. (2015) recommend using their triceps and subscapular skinfolds reference values for Hispanic American children.

Increasingly, reference data based on anthropometric measures of adiposity based on skinfolds are becoming available from low and middle-income countries. Khadilkar et al. (2015) have published reference percentiles for triceps skinfold thickness for Indian children aged 5-17y, whereas Pandey et al. (2008) provide percentiles for both triceps and subscapular skinfolds for urban Asian Indians aged 14-18y. Again, these percentiles differed and were higher than those recorded for U.S. children. Even infants in South Asia appear to have subscapular skinfolds at birth that are higher than those for comparable birthweight Caucasian babies, despite having other body measurements that are smaller (Yajnik et al., 2003).

Age- and sex-standardized percentile reference curves for triceps and subscapular skinfold thicknesses are especially useful in remote emergency settings, in bed-bound hospitalized patients, and when other medical conditions are present that preclude the evaluation of weight, height, and body composition (Heymsfield & Stevens, 2017).

11.1.2 Assessing body fat with skinfolds

Skinfold measurements at a single or multiple sites can be used to estimate total body fat or percentage body fat. Calculation of percentage body fat is based on the assumption that fat mass is adjusted for body weight, even though percentage body fat is not fully independent of body size (Wells, 2014). Furthermore, high values for percentage body fat might reflect either high fat mass or low fat-free mass, as noted earlier (Wells, 2019).

If a single skinfold measurement approach is used, it is critical to select the skinfold site that is most representative of the whole subcutaneous fat layer, because subcutaneous fat is not uniformly distributed about the body. Unfortunately, the most representative site is not the same for both sexes, nor is it the same for all ages, ethnicities, or degree of adiposity. Hence, it is not surprising that there is no general agreement as to the best single skinfold site as an index of total body fat. In the past, the triceps skinfold thickness has been the site most frequently selected by nutritionists for a single, indirect estimate of body fat.

To account for the differing distribution of sub&sny;cutaneous fat, investigators often recommend taking one limb skinfold (right triceps) and one body skinfold measurement (right subscapular). For example, persons of African descent tend to have less subcutaneous fat in the extremities than in the trunk relative to Caucasians, irrespective of age and athletic status (Wagner & Heyward, 2000).

More than 100 formulae have been developed to estimate percentage body fat from skinfold thickness measurements alone. The formulae have been established across varying populations, using numerous protocols with deviations in the skinfold sites measured (Lohman et al, 1988). Unfortunately, discrepancies have been reported when different formulae are applied on the same set of individuals. This finding has led to the proposal that the sum of skinfold sites (in mm) (preferably using eight sites) may provide a more accurate and reliable outcome of body fat than using an indirect method based on anthropometric-based prediction formulae (Kasper et al., 2021).

The measurement of multiple skinfolds and not just a single skinfold to estimate body fat is particularly advisable when individuals are undergoing rapid and pronounced weight gain. Changes in the energy balance are known to alter the rate of fat accumulation differently among skinfold sites(Heymsfield et al., 1984)

11.1.3 Body adiposity index

The body adiposity index (BAI) is a surrogate measure of adiposity which is calculated as:

$$\text{BAI } (\%\text{fat}) = \frac{\text{Hip circumference}}{(\text{ Height })^{1.5}} - 18$$

The body adiposity index (BAI) was developed by Bergman et al. (2011) in part as a result of the inability of BMI to distinguish between fat and fat-free mass. Several studies of adults have reported positive correlations of BAI with BMI and waist circumference (Nickerson et al., 2015). Further, BAI has been validated as a measure of percentage body fat against both DXA (Johnson et al., 2012; Sun et al., 2021; Nickerson et al., 2015), and more recently the 4-component model (Fedewa et al., 2019).

The 4-component model is considered the gold standard criterion method for measuring percentage body fat because the technique reduces the need for theoretical assumptions when calculating body composition outputs; information on body weight, body volume, total body water (TBW) and bone mineral mass are each collected separately (Wells, 2014). Nevertheless, results on the relative accuracy of BAI as a surrogate indicator of percentage body fat have been inconsistent, and appear to be dependent on sex (Johnson et al., 2012; Fedewa et al., 2019), race-ethnicity (Johnson et al., 2012; Ramírez-Vélez et al., 2016), level of adiposity (Bergman et al., 2011; Johnson et al., 2012; Sun et al., 2021) and activity level (Esco, 2013). More research employing prospective studies are needed to establish the predictive ability of BAI for various health outcomes in differing population groups.

11.1.4 Arm-fat area

The calculated cross-sectional area of arm fat, derived from skinfold thickness and arm circumference measurements, has been used as an index of total body fat, especially in emergency settings. Arm-fat area correlates more significantly with total body fat (i.e., fat weight) than does a single skinfold thickness at the same site. In contrast, the estimation of percentage of body fat from arm fat area is no better than the corresponding estimation from the skinfold measurement, particularly in males (Himes et al., 1980).

The advantage of using arm-fat area to estimate body fat is expected; more fat is needed to cover a large arm with a given thickness of subcutaneous fat than to cover a smaller arm with the same thickness of fat. Subcutaneous fat, however, is not evenly distributed around the limbs or trunk. For example, triceps skinfolds are consistently larger than the corresponding biceps skinfold, and, as a result, either the sum or the average of these should theoretically be used for the calculation of mid-upper-arm-fat area (Himes et al., 1980). In practice, however, limb fat area reference data are only available based on triceps skinfold and mid-upper-arm circumference (MUAC) measurements (Frisancho, 1990; Addo et al., 2017). Trained examiners using standardized techniques should be used for these measurements; for details see Sections 11.1.1 and 11.2.1.

Calculation of arm-fat area

The equation for calculating mid-upper-arm-fat area is:

$$\mathrm{AFA} = (\mathrm{SKF} \times \mathrm{MUAC}/2) - \left(\pi \times (\mathrm{SKF})^2/4\right)$$

where AFA = mid-upper-arm-fat area (mm^2), MUAC = mid-upper-arm circumference (mm), and SKF = triceps skinfold thickness (mm). This equation is based on several assumptions, each of which may result in inaccuracies, leading to an underestimate of the degree of adiposity (Rolland-Cachera et al., 1997). The equation assumes that the limb is cylindrical, with fat evenly distributed about its circumference, and also makes no allowance for variable skinfold compressibility. This compressibility probably varies with age, sex, and site of the measurement, as well as among individuals, and is a source of error in population studies when equal compressibility of skinfolds is assumed.

Arm-fat areas calculated from this equation were reported to agree within 10% to values measured by computerized axial tomography on normal weight adults. However, for obese subjects, differences were greater than 50% (Heymsfield et al., 1982). A correction for skinfold compressibility may be advisable in future studies.

A simplified index has also been proposed, the upper-arm-fat estimate (UFE in cm^2)

$$\mathrm{UFE} = \mathrm{MUAC} \times \mathrm{TSF}/2$$

This equation was validated by comparing arm-fat areas assessed by magnetic resonance imaging (MRI) and anthropometry in 11 obese and 17 control children. Both the traditional upper-arm-fat area and the upper-arm-fat estimate (UFE) were calculated. Results indicated that the UFE measurements were close to the MRI estimates (Rolland-Cachera et al., 1997). Nevertheless, this simplified index has had limited use.

Interpretive criteria

Reference data for mid-upper-arm-fat area were compiled by Frisancho (1981) from the earlier NHANES I survey (1971-1974) when smoothing techniques were not applied. More recently, Addo et al. (2017) have derived sex-specific percentile curves for arm fat area based on a wider age range of U.S children (i.e., 1-20y) who were also included in the development of the CDC 2000 BMI growth charts (Kuczmarski et al., 2000a). For these charts, the measurements for all children between 1963 and 1994 were included, with the exception of those > 6y of age who were measured between 1988 and 1994. These children were excluded because of the rising prevalence of obesity (Kuczmarski et al., 2002). Figure 11.6 presents the arm-fat area percentiles for female US children and adolescents aged 1-20y.

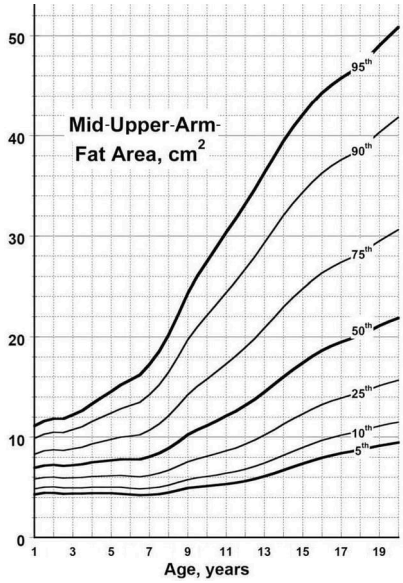


Figure 11.6 Upper-arm-fat-area-for-age percentiles for female US children and adolescents aged 1-20y. Redrawn from Addo et al. (2017).

In addition to age and sex, height was also found to influence the ranking of mid-arm-fat area in this study. Hence, prediction equations for height-for-age adjusted Z-scores for arm-fat area-for-age are also reported for males and females; see Addo et al. (2017) for more details. However, there is evidence that population-specific reference data are needed for arm-fat area. Oyhenart et al. (2019a) compared the arm-fat area values of children aged 4-14y in Argentina with the U.S. reference data (Addo et al., 2017). The higher mean values of arm-fat area for 3rd, 50th, and 97th percentiles were indicative of greater adipose tissue for Argentinian boys and girls than for comparably aged U.S. children.

11.1.5 Calculation of body fat from anthropometric variables via body density

Measurements of anthropometric variables from multiple anatomical sites, including skinfolds, are also used to estimate body density from which the percentage of body fat, and subsequently total body fat are calculated. The method involves:

- Determination of appropriate skinfolds and other anthropometric measurements for the prediction of body density; the selection of the sites depends on the age, sex, ethnicity/race, and population group under investigation
- Calculation of body density, using an appropriate prediction equation
- Calculation of percentage of body fat from body density using an empirical densitometric equation
- Calculation of total body fat and/or the fat-free mass:

$$\text{Total body fat (kg)} = \text{ body weight (kg)} \times \%\text{ body fat } /100$$
$$\text{Fat-free mass (kg)} = \text{ body weight (kg)} - \text{ total body fat (kg)}$$

Choice of appropriate anthropometric variables to estimate body density (Steps 1 and 2)

Many studies have investigated the best combination of skinfolds and other anthropometric measurements from which to derive a regression equation for the initial estimation of body density (Steps 1-2), prior to the calculation of percentage body fat (Step 3) by applying the empirical equations of Siri (1956) or Brožek (1963).

Numerous prediction equations are available to estimate body density from anthropometric variables for population groups ranging from sedentary to athletic and from children to the elderly (Provyn et al., 2011). Rarely have the studies recommended the same combination of measurements. The selection of the most appropriate prediction equation should be based on the characteristics of the population on which the chosen equation was originally validated.

Several studies have examined the predictive accuracy and the applicability of the prediction formulae available for estimating body density and subsequently percentage body fat. For example, Provyn et al. (2011) reported that even when the chosen predictive formulae were matched to the characteristics of the population (e.g., age, gender, ethnicity, activity level) on which the equation was originally validated, the prediction formulae investigated (n=57) were not necessarily reliable tools for predicting percentage body fat in Caucasian adults. This conclusion was based on the percentage body fat estimates generated from dual-energy X-ray absorptiometry (DXA), and confirmed earlier from body fat data (in g) generated from direct dissection of domestic porcine hind legs (Provyn et al., 2008). Hence, the application of these prediction formulae for estimating percentage body fat on age-matched, apparently healthy individuals remains questionable.

Certainly, these prediction formulae should not be used to predict body density in undernourished individuals, as there is a decreasing correlation between skinfold thickness and total body fat content with increasing severity of undernutrition. This change in correlation may arise from a shift of fat storage from the regions represented by the subscapular and triceps skinfolds to other subcutaneous sites. Alternatively, a shift from subcutaneous to deep visceral sites may occur (Spurr et al., 1981).

Calculation of percentage body fat from body density using empirical equations (Step 3)

In most cases, the final stage in the calculation of the percentage of body fat (F) from measurements of skinfolds and other anthropometric variables is the selection of an empirical densitometric equation relating fat content to body density (D). Several densitometric equations have been derived based on the two-component model for body composition in which body weight is divided into fat and fat-free mass, relying on assumptions that ignore inter-individual variability in the composition of fat-free mass. All the classical densitometric equations assume: (a) the density of the fat-free mass is relatively constant; (b) the density of fat for normal persons does not vary among individuals; (c) the water content of the fat-free mass is constant; and (d) the proportion of bone mineral (i.e., skeleton) to muscle in the fat-free body is constant. All authors used the equation:

$$\%\ \mathrm{F} = ((\mathrm{C}_1/\mathrm{D}) - \mathrm{C}_2) \times 100\%$$

but different authors used different values for the density of fat and the fat-free mass and as a result the values for C_1 and C_2 differ slightly:

$$C_1 = 4.950, C_2 = 4.500(\text{ Siri, 1961 })$$
$$C_1 = 4.570, C_2 = 4.142(\text{ Brožek et al. 1963 })$$
$$C_1 = 5.548, C_2 = 5.044\ (\text{Rathburn \& Pace, 1945})$$

All assume the density of fat and the fat-free mass are constant by age and sex. Siri (1961) assumed that the densities of fat and the fat-free mass are 0.90 and 1.10kg/L respectively. Brožek et al. (1963) and Rathburn & Pace (1945) used the concept of a reference man of a specified density and composition. These equations came from the chemical analysis of a few adult cadaver dissections, animal data, and indirect estimates of fat-free mass in human subjects (Siri, 1961; Brožek et al., 1963; Heymsfield et al., 1991).

None of these classical empirical equations relating fat content to body density, however, are suitable in adult patients in whom the composition of fat-free mass may be abnormal. This will include patients undergoing hyperalimentation with high-sodium fluids, or with congestive heart failure or liver disease, as total body water content as a fraction of fat-free mass may be markedly

higher in these patients, thus violating the assumption that the water content of the fat-free mass is constant (Heymsfield & Casper, 1987). In these circumstances, the density of fat-free mass is decreased. Not surprisingly, in patients with diseases associated with under-mineralisation, the density of fat-free mass is also decreased. Consequently, in all these patients, fatness will be overestimated (Wells & Fewtrell, 2006).

More recent research has raised concerns over the assumption of constant properties for hydration and density of fat-free mass when these classical empirical equations are applied to assess body composition not only in patients with certain diseases, but also in healthy children and adolescents, the elderly, and those with obesity. Although fat has relatively uniform properties throughout the life course (zero water and a density of 0.9007kg/L), fat-free mass, in contrast, has different properties in children compared to adults. This arises because of chemical maturation of the fat-free mass during growth which results in higher levels of water and lower levels of mineral and proteins. Nevertheless, the adult-derived values for the density and hydration of fat-free mass and applied in the classical equations have often been used to study body composition in children.

In an effort to improve the accuracy in the estimates of percentage body fat in children and adolescents based on the two-component model, Wells et al. (1999) measured the density and hydration of fat-free mass in children (n=41) aged 8-12y using the 4-component model which divides body weight into fat, mineral, and protein and overcomes the limitations associated with the assumptions of constant properties for hydration and fat-free mass density (Table 11.5).

Table 11.5 Median values for males for hydration, density, and constants (C_1 and C_2) for the paediatric version of Siri's equation, obta using the LMS (lambda-mu-sigma) method. Data from Wells et al. (2010) who also present comparable data for females.

Age	Hydration (%)	Density (kg/L)	C_1	C_2
5	76.5	1.0827	5.36	4.95
6	76.3	1.0844	5.32	4.90
7	76.1	1.0861	5.28	4.86
8	75.9	1.0877	5.24	4.82
9	75.7	1.0889	5.21	4.79
10	75.5	1.0900	5.19	4.76
11	75.3	1.0911	5.16	4.73
12	75.2	1.0917	5.15	4.72
13	75.0	1.0920	5.14	4.71
14	74.8	1.0927	5.13	4.69
15	74.4	1.0942	5.09	4.66
16	74.0	1.0960	5.05	4.61
17	73.7	1.0978	5.02	4.57
18	73.5	1.0991	4.99	4.54
19	73.4	1.1000	4.97	4.52
20	73.3	1.1006	4.96	4.51

They reported the measured fat-free mass density for the children to be significantly lower than the adult value (1.0864kg/L vs. 1.1kg/L), whereas that for the measured fat-free-mass hydration was higher (75.3% vs. 73.2%). In a later study comprising a larger sample of children (n=533) and wider age range (4-23y), Wells et al. (2010) developed empirical reference data for density and hydration of fat-free mass for children from age 5-20y based on the 4-component model (i.e., body weight, total body water, bone mineral content, and body volume). Table 11.5 presents the median values for hydration, density, and constants (C_1 and C_2). In addition, they developed prediction equations for the density and hydration of the fat-free mass based on age, sex, and body mass index standard deviation score (BMI SDS) using their 4-component measurements of body composition; see Wells et al. (2010) for more details.

Note that the values for C_1 and C_2 constants for the adult males (age 20y) shown here are similar to the corresponding values shown in the Siri equation above (i.e., 4.95 for C_1 and 4.50 for C_2), whereas for adult females, the corresponding values are slightly lower: 4.90 for C_1 and 4.44 for C_2 at 20 y. With the substitution of the age- and sex-specific $C_1$and C_2 constants in Table 11.5 for the C_1 (4.95) and C_2 (4.50) constants in the Siri equation, the accuracy of the two-component model for estimating fat mass of a healthy pediatric population could be improved.

More recent research indicates that nutritional status should also be considered when selecting values for both the density and hydration of fat-free mass using a two-component model for body composition. Gutierrez-Martin et al. (2019) reported increasing values for hydration but decreasing values for the density of fat-free mass in the children with heavier BMIs (Figure 11.7), a trend that has been observed earlier in children (Haroun et al., 2005) and adults (Waki et al., 1991).

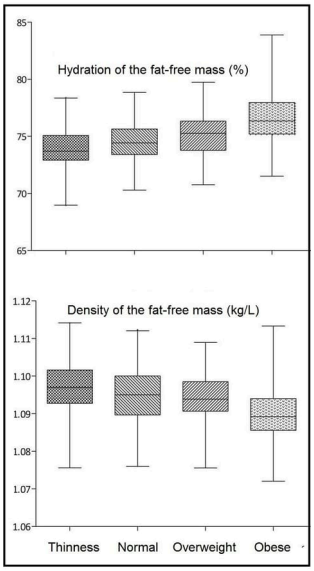


Figure 11.7 Values for hydration and density of fat free mass based on the four-component model and stratified by nutritional status grouped by BMI SD score for UK subjects aged 4-22y. Redrawn and abbreviated from Gutierrez-Marin et al. (2019).

Consequently, these investigators developed a method whereby corrections for the density of fat-free mass could be made for children with obesity and thus improve the accuracy of the two-component model for estimating fat mass in obese children. For more details of the adjustment process, see Gutierrez-Martin et al. (2021).

Similar trends in the values for the hydration and density of fat-free mass compared to the classic adult values applied in the Siri densitometric equation have been observed among older adults aged > 60y; values for fat-free mass density were lower but higher for the hydration fraction. These measurements were reported in studies of both Hispanic Americans (Gonzalez-Arellanes et al., 2019). and obese Mexicans`(González-Arellanes et al., 2021) aged > 60y and were based on the 4-component model. Their findings also suggest that modifying the assumptions regarding both the density and hydration values for fat-free mass applied in the classical densitometric empirical equations may also be appropriate for the elderly and in conditions of obesity.

Recognition for the need to modify these classical empirical equations which assume constant properties of fat-free mass (hydration and density) has led to an increase in the measurement of body composition *in vivo* using the 4-component model. This method is considered the gold standard for measuring body composition because it reduces the need for theoretical assumptions. See Chapter 14 for more discussion of the *in vivo* methods used to measure body composition.

11.1.6 Waist-hip circumference ratio

The waist-hip circumference ratio (waist circumference divided by hip circumference) (WHR) is a simple method for distinguishing between fatness in the lower trunk (hip and buttocks) and fatness in the upper trunk (waist and abdomen areas). Lower trunk fatness (i.e., lower waist to hip ratio) is often referred to as “gynoid obesity” because it is more typical of females. Upper trunk or

central fatness (higher waist to hip ratio) is called “android obesity” and is more characteristic of males. Nevertheless, obese men and women can be, and often are, classified into either group.

The fat depots assessed by the WHR are mainly subcutaneous (external or outer) and visceral (internal or deep). Use of the WHR rose dramatically following several reports confirming that WHR separately or in combination with BMI was associated with increased risk of death, coronary heart disease and type 2 diabetes mellitus (Krotkiewski et al., 1983; Larsson et al., 1984).

The application of new laboratory methods including computer tomography and magnetic resonance imaging has led to semi-quantitative estimates of the total fat stored within the abdomen (i.e., intra-abdominal fat). Ashwell et al. (1985) were the first investigators to show highly significant correlations between intra-abdominal fat (visceral adipose tissue) and the ratio of waist-to-hip circumference. Their findings led to the proposal that the metabolic complications of obesity shown to be associated with a high WHR, may be related specifically to the amount of intra-abdominal (visceral) fat. Recently, in a meta-analysis of 21 prospective cohort studies in which waist-hip ratio was measured as an indicator of abdominal obesity, the risk of cardiovascular disease rose continually with the increase in WHR when they exceeded a certain range (Xue et al., 2021). Based on these results the investigators advised that men should keep their WHR below 0.9 to maintain cardiovascular fitness, whereas women should keep their WHR as small as possible within the normal range.

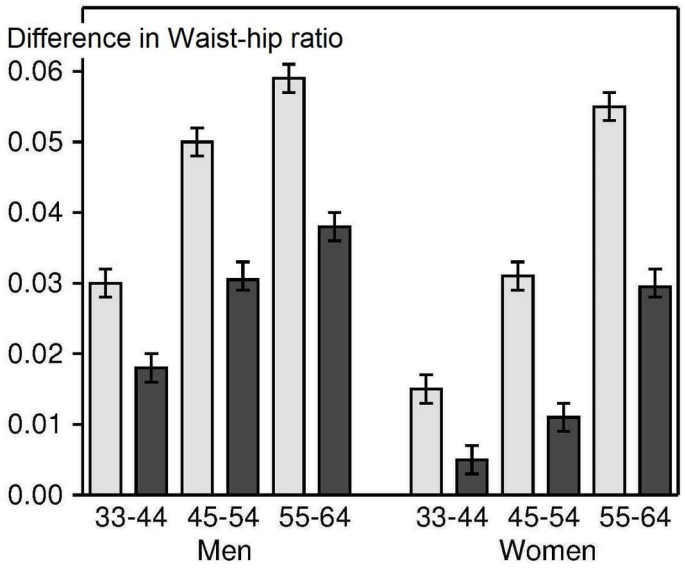


Figure 11.8. Effect of age on waist-hip ratio (WHR) shown by the difference in WHR in older subjects relative to the WHR at age 25-34y. Pooled data from 19 male and 18 female populations in the second MONitoring trends and determinants in CArdiovascular disease (MONICA) survey. Unadjusted differences (GREY - adjusted only for population) and adjusted differences (BLACK - adjusted for population, height and body mass (BMI). Redrawn from Molarius et al. (1999).

Several studies in adults have shown that the WHR varies with age and the degree of overweight, in addition to sex (Stevens et al., 2010). Jones et al. (1986) measured the WHR of a semi-random, age-stratified sample of 4349 British Caucasian men 20-64y. They noted that the ratio increased with both age (curvilinearly) and excessive weight. In the WHO multinational MONItoring of trends and determinants of CArdiovascular disease (MONICA) project (Molarius et al., 1999), the WHR was also reported to increase with age in both men and women (Figure 11.8), and to be higher in men than in women.

There is also some evidence that WHR varies with ethnicity. In the MONICA project, a standard protocol was used to measure waist and hip circumference in men and women age 25-64y in 19 countries. Mean waist-hip ratio varied considerably among the study population, ranging from 0.87-0.99 for men and from 0.76-0.84 for women. To date, most of the evidence for race/ethnic differences relates to Asian adults in whom lower WHRs have been associated with an increased metabolic risk compared to Europeans, probably because of higher body fat and visceral adipose tissue (Lear et al., 2010).

Relationships between the WHR and age, sex, and race/ethnicity have also been investigated in children. In the U.S. NHANES III,

mean WHR varied consistently with age, sex, and ethnic group in children and adolescents aged 4-19y, as shown in Figure 11.9. Ratios were highest in Mexican American boys (Gillum, 1999).

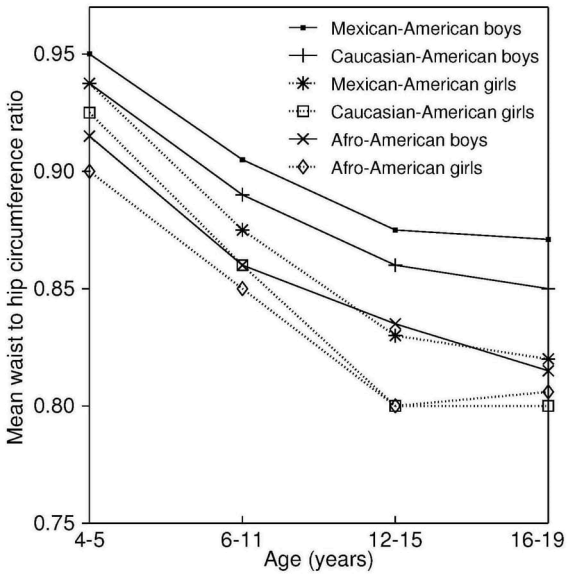


Figure 11.9. Mean waist-to-hip circumference ratio in male and female children and young adults of three racial groups. NHANES III data (1988-1994). Redrawn from Gillum (1999).

A variety of adverse health outcome measures have been examined in relation to WHR, most of which have been based on cross-sectional studies using differing methods to measure WHR. Consequently, comparisons across studies are difficult, as emphasized by Lear et al. (2010). However, in a prospective cohort study involving 15062 participants from Norfolk, U.K., WHR appeared to have the best predictive value for cardiovascular disease and mortality compared with BMI and percentage body fat (Myint et al., 2014). In some prospective studies, WHR has been associated with a higher risk for all-cause and cardiovascular mortality particularly in women (Rost et al., 2018).

Measurement of waist-hip ratio

WHO (2011) has recommended standardized protocols for the measurements of waist and hip circumference for international use. WHO (2011)considered the following elements when developing the protocols: anatomical placement of the measuring tape, its tightness, and the type of tape used; the subject's posture, phase of respiration, abdominal tension, stomach contents, and clothing.

After an extensive review, WHO (2011) concluded that waist circumference should be measured at the midpoint between the tenth rib (i.e., the lowest rib margin) and the top of the iliac crest, using a stretch-resistant tape that provides a constant 100g tension. Hip circumference should be measured around the widest portion of the buttocks, with the tape parallel to the floor.

To perform the waist-circumference measurement, the lowest rib margin is first located and marked with a felt tip pen. The iliac crest is then palpated in the midaxillary line and the top of the iliac crest is also marked. An elastic tape can then be applied horizontally at the mid-point between the lowest rib margin and the highest point of the iliac crest: it is tied firmly so that it stays in position around the abdomen about the level of the umbilicus. The elastic tape thus defines the level of the waist circumference, which can then be measured by positioning the stretch-resistant tape over the elastic tape (Jones et al., 1986). Alternatively, a washable marker can be used to landmark the location of the tape. The stretch-resistant tape used for the measurement should provide a constant 100g tension. This can be achieved through the use of a special indicator buckle that reduces differences in tightness.

The subject should wear little clothing and be asked to stand erect with feet close together, arms at the side, with their body weight evenly distributed across the feet. The subject should be relaxed and asked to take a few deep, natural breaths. The measurement should be taken at the end of a normal expiration to prevent the subject from contracting their muscles or from holding their breath. The measurement is taken when the tape is parallel to the floor, and the tape is snug, but does not compress

the skin. The reading is taken to the nearest millimeter. Each measurement should be repeated twice. If the two measurements are within 1cm of one another, the average should be calculated, but if the difference exceeds 1cm, then the two measurements should be repeated.

For the hip circumference measurement, the subject should stand erect with arms at the side and feet together, with body weight equally distributed across the feet. The measurement should be taken with the stretch-resistant tape used for the waist circumference measurement at the point yielding the maximum circumference over the buttocks. The tape must be held parallel to the floor, touching the skin but not indenting the soft tissue. The measurement is taken to the nearest millimeter. Again, each measurement should be taken twice. If the two measurements are within 1cm of one another, the average should be calculated, but if the difference exceeds 1cm, then the two measurements should be repeated. The degree to which factors such as postprandial status, standing position, and depth of inspiration contribute to error in the measurement of waist-hip circumference ratio is uncertain.

Interpretive criteria

Bjôrntorp (1987) was the first to suggest that waist-hip ratios > 1.0 for men and > 0.85 for women indicated abdominal fat accumulation and an increased risk of cardiovascular complications and related deaths. Subsequently, many countries and settings have identified sex-specific cutoff points for WHRs, some also recommending ethnically based cutoff points particularly for populations of Asian descent. Generally, most of the cutoffs chosen have been based on disease risk (e.g., cardiovascular disease, type 2 diabetes and risk factors of cardiovascular disease) and on hard outcomes such as mortality. Japan is an exception as their cutoffs are based on assessment of visceral adipose tissue from computerized tomography, and hence on the extent to which measurements predict intra-abdominal fat rather than disease risk (WHO, 2011).

Currently, WHO (2011) define abdominal obesity and risk of metabolic consequences as a WHR > 0.90 for men and WHR > 0.85 for women. The cutoffs recommended by the U.S. Department of Health and Human Services for WHR are > 0.95 for men and > 0.80 for women.

WHO (2011) emphasize that further studies are needed to establish whether cutoff points for WHRs should be specific to age and ethnicity, given the known ethnic variations in body fat distribution, especially in populations of Asian origin (Wagner & Heyward, 2000; Lear et al., 2010). Further, different contributions of muscle mass and bone structure, as well as stature and abdominal muscle tone, may all lead to different associations between WHR and abdominal fat accumulation.

Figure 11.10

Figure 11.10 Visceral adipose tissue area and liver attenuation by WC tertiles within each BMI category in men. Redrawn from Nazare et al. (2015), who also present data for women.

The validity of serial measurements of WHR to measure changes in intra-abdominal visceral fat over time is uncertain. For example, any beneficial reductions in abdominal fat will not be evident when a ratio such as WHR is used if both the numerator and denominator values change in response to treatment. Consequently, waist circumference alone is now the preferred index for monitoring loss of visceral adipose tissue, and is discussed below.

11.1.7 Waist circumference

Studies have shown that compared with the WHR, waist circumference alone is more strongly associated with the amount of intra-abdominal fat (i.e., visceral fat tissue) (Snijder et al., 2006; Neeland et al., 2019). Moreover, with an increase in waist circumference, there is a corresponding increase of visceral adipose tissue, the fat depot known to convey the strongest health risk. For example, in a large study in 29 countries waist circumference and BMI were measured, and visceral adipose tissue assessed directly using computer tomography (Nazare et al., 2015). A global cardiovascular risk score was also calculated from the sum of eight individual risk factor subscores based on a series of clinical biomarkers, all measured in one laboratory. As shown in Figure 11.10, visceral adipose tissue increased significantly while liver attenuation (inversely correlated with liver fat, a depot of ectopic fat) decreased significantly across the waist circumference tertiles, within each of the three BMI categories. Further, the measured cardiometabolic risk score reflecting the number of cardiometabolic abnormalities, was significantly correlated to visceral adipose tissue, waist circumference, and BMI in men and women.

Table 11.6 Cardiometabolic risk score (CMR score) values across tertiles of waist circumference (WC) within each of the 3 body mass index (BMI) categories
**p < 0.05, **p < 0.01, ***p < 0.0001, denote significantly different from the first WC tertile group within the same BMI category and*
†p < 0.05, ††p < 0.01, †††p < 0.0001 denote significantly different from the middle WC tertile group within the same BMI category.
T1, T2 and T3 are the WC tertile groups. All statistical analyses were adjusted for age, ethnicity, physician's specialty, smoking status and educational level. BMI = body mass index; CMR = cardiometabolic risk; WC = waist circumference. Data are CMR score means ±SEM. Data from Nazare et al. (2015).

	WC tertiles		
	T1	T2	T3
Men - BMI>			
< 25kg/m^2	WC ≤ 84cm 2.1 ± 0.1	84 < WC ≤ 90cm 2.5 ± 0.1**	WC > 90cm 2.7 ± 0.1***
25kg/m^2 to < 30kg/m^2	WC ≤ 95cm 2.7 ± 0.1	95 < WC ≤ 101cm 3.3 ± 0.1**	WC > 101cm 3.6 ± 0.1***
≥ 30kg/m^2	WC ≤ 108cm 3.7 ± 0.1	108 < WC ≤ 116cm 4.1 ± 0.1*	WC > 116cm 4.5 ± 0.1 **†
Women - BMI>			
< 25kg/m^2	WC ≤ 76cm 1.5 ± 0.1	76 < WC ≤ 83cm 1.9 ± 0.1**	WC > 83cm 2.7 ± 0.1***†††
25kg/m^2 to < 30kg/m^2	WC ≤ 87cm 2.5 ± 0.1	87 < WC ≤ 93cm 3.3 ± 0.1***	WC > 93cm 3.8 ± 0.1***††
≥ 30kg/m^2	WC ≤ 100cm 3.4 ± 0.1	100 < WC ≤ 108cm 3.8 ± 0.1	WC > 108cm 4.6 ± 0.1**††

Table 11.6 presents a comparison of the cardiometabolic risk scores by waist circumference tertile groups in each of the three BMI categories. Note the increase in cardiometabolic risk score for both males and females in all three categories of BMI across the waist circumference tertile groups.

Data from numerous other epidemiological studies have also shown that visceral adipose tissue is an independent risk marker of cardiovascular and metabolic morbidity and mortality. A study by Hiuge-Shimizu et al. (2012) on visceral fat accumulation, measured on computer tomography scans in 12,443 Japanese subjects, indicated that an absolute visceral fat area of about 100cm^2 equated with an increased risk of factors associated with obesity-related cardiovascular morbidity. Moreover, this relationship was irrespective of gender, as shown in Figure 11.11, as well as age and BMI. The obesity-related cardiovascular risk factors assessed in this study were hyperglycemia, dyslipidemia, and elevated blood pressure.

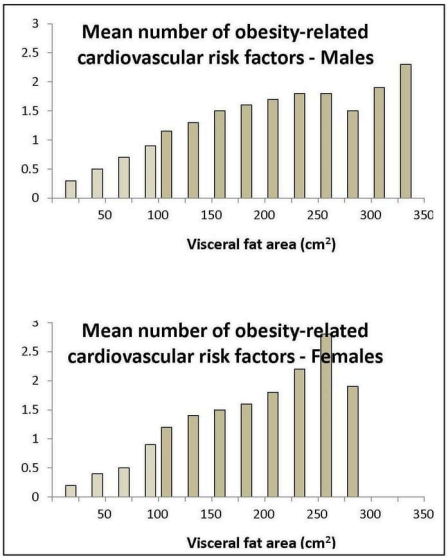


Figure 11.11 Association between the mean visceral fat area (VFA) and obesity-related cardiovascular risk factors. Pale bars: subjects with less than 1.0 risk factors; darker bars: subjects with more than 1.0 risk factors. Redrawn from Hiuge-Shimizu et al.

These findings have highlighted the importance of measuring waist circumference along with BMI in clinical practice to assess the distribution of visceral (intra-abdominal) adipose tissue between individuals (Ross et al., 2020). Moreover, hip circumference is more difficult to measure than waist circumference. (2012).

Listed below are the health and metabolic abnormalities that have been associated with an excess deposition of visceral adipose tissue and ectopic fat irrespective of total adiposity estimated by BMI (Neeland et al., 2019).

- Insulin resistance
- Impaired glucose tolerance
- Type 2 diabetes
- Cardiovascular disease
  - Hypertension
  - Heart failure
  - Coronary heart disease, myocardial infarctions
  - Valve diseases
  - Arrhythmias
- Respiratory diseases
  - Sleep apnoea
  - Chronic obstructive pulmonary disease
- Brain health
  - Stroke, necrosis
  - Reduced brain size
  - Reduced grey matter
  - Reduced cognitive function
  - Dementia
- Cancers
- Others
  - Reduced bone density
  - Polycystic ovary syndrome
  - HIV infection and antiretroviral therapy as both can contribute to the accumulation of visceral adipose tissue and ectopic fat.

Note, however, that the evidence for a causal relation with some of these conditions is insufficient. For more details of the constellation of metabolic abnormalities associated with an excess of visceral adipose tissue, see Neeland et al. (2018).

More recently, with the development of medical imaging, the detection and measurement of fat in areas of the body where fat is not physiologically stored has also been made. These studies have shown that at any given BMI, excess visceral adiposity is often associated with an increased accumulation of fat in normally lean tissues such as the liver, pancreas, heart, and skeletal muscle, a condition termed ectopic fat deposition. As noted earlier, emerging evidence suggests that the deposition of ectopic fat might contribute to increased risk of atherosclerosis and cardiometabolic risk (Neeland et al., 2019).

The causal mechanisms whereby an excess of visceral adipose tissue is related to the cardiometabolic complications are not yet fully established. Three mutually exclusive scenarios have been proposed, and are reviewed by Neeland et al. (2019):

- Visceral adipose tissue has metabolic properties that are distinct from subcutaneous adipose tissue;
- Excess visceral adipose tissue induces inflammation;
- Visceral adipose tissue is a marker of increased ectopic fat deposition (including hepatic and epicardial fat)

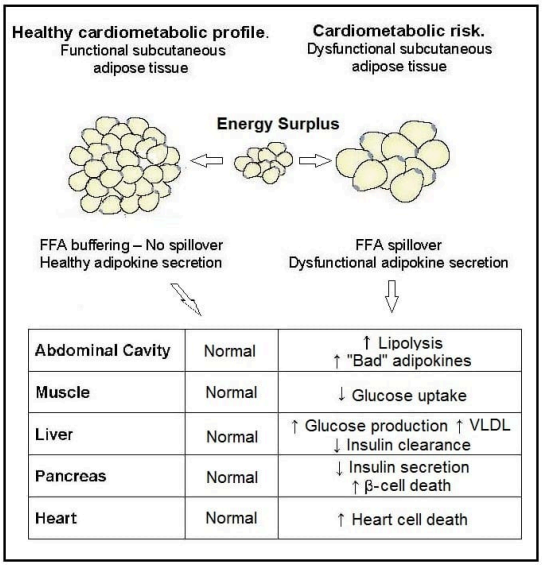


Figure 11.12 Overview of potential role of functional and dysfunctional adipose tissue contributing to increased cardiometabolic risk. FFA = free fatty acid; VLDL = very-low-density lipoprotein cholesterol. Redrawn from Ross et al. (2020).

An overview of the potential role of functional and dysfunctional adipose tissue contributing to increased cardiometabolic risk is presented in Figure 11.12. In a healthy cardiometabolic profile, the ability of subcutaneous adipose tissue to expand through hyperplasia (generation of new fat cells) allows the safe storage of the excess energy from the diet into a properly expanding subcutaneous 'metabolic sink'. When this process becomes saturated or in a situation where adipose tissue has a limited ability to expand, there is a spillover of the excess energy, which must be stored in visceral adipose tissue as well as in normally lean organs such as the skeletal muscle, the liver, the pancreas, and the heart, a process described as ectopic fat deposition. Visceral adiposity is associated with a hyperlipolytic state resistant to the effect of insulin along with an altered secretion of adipokines including inflammatory cytokines, whereas a set of metabolic dysfuntions are specifically associated with increased skeletal muscle, liver, pancreas, and epicardial, pericardial, and intra-myocardial fat. For more discussion, see the consensus documents by the International Atherosclerosis Society (IAS) and International Chair of Cardiometabolic Risk (ICCR) Working Group on Visceral Obesity (Neeland et al., 2019; Ross et al., 2020).

Neeland et al. (2019) have also reviewed the response of visceral and ectopic fat to treatment. Briefly, both exercise and dietary interventions are reportedly associated with a substantial reduction in visceral adipose tissue independent of age, sex, and ethnic origin, and irrespective of amount or intensity of exercise. Moreover, randomized controlled trials that have reported lifestyle-induced reductions in visceral adipose tissue and thus waist circumference have also shown they are associated with improvements in cardiometabolic risk factors with or without corresponding weight loss.

These observations, taken together, emphasize the importance of developing simple clinically applicable tools, previously validated with imaging data, with the ability to monitor changes in visceral and ectopic fat over time (Neeland et al., 2019; Ross et al., 2020). In this way, the definition of high-risk overweight and obesity could be refined. In the meantime, Neeland et al. (2019) suggest that the addition of the measurement of plasma triglyceride concentrations to the measurements of waist circumference

may be helpful as a screening tool to identify individuals likely to be characterized by the cluster of abnormalities of the metabolic syndrome, as long as validated waist circumference cutoff values are applied.

Measurement of waist circumference

A consensus on the optimal protocol for the measurement of waist circumference has not yet been reached. Currently two sites are used: (a) at the natural waist, i.e., mid-way between the tenth rib (the lowest rib margin) and the iliac crest (i.e., the superior border of the wing of the ilium), as proposed by WHO (2011) and (b) at the umbilicus level (van der Kooy & Seidell, 1993). In the future, adopting a standard approach by using the protocol described by WHO (2011) and described in Section 11.1.7, is recommended. In this way differences that might exist in absolute waist circumference measurements due to the difference in protocols will be avoided (Ross et al., 2020).

Interpretive criteria

Waist circumference cutoffs in adults have been developed as simple surrogate markers to identify the increased risk associated with excess visceral adipose tissue (intra-abdominal fat). Consequently, measurements of waist circumference should be included routinely along with BMI by health practitioners in the evaluation and management of patients with overweight and obesity (Ross et al., 2020).

In several countries a single cutoff threshold for white adults (> 102cm for men and > 88cm for women) is currently used to denote a high waist circumference, irrespective of BMI category (Molarius et al., 1999; Health Canada, 2003). These same sex-specific cutoffs have been proposed by WHO (2011). They were based on cross-sectional data in Caucasian adults in whom the specified sex-specific waist circumference cutoffs corresponded to a BMI of 30.0kg/m^2, the BMI cutoff designated for obesity. Hence, they were not developed based on the relationship between waist circumference and adverse health risk (Ross et al., 2020).

WHO (2011) recognized that population-specific cutoffs may be warranted in view of differences in the level of risk associated with a particular cutoff across populations, depending on levels of obesity and other risk factors for cardiovascular disease and type 2 diabetes. However, they emphasize that further prospective studies using representative populations are needed to understand the genetic and lifestyle factors that may be contributing to the reported regional variations in waist circumference (Lear et al., 2010). Consequently, to date, WHO (2011) have not recommended ethnicity-specific cutoffs for waist circumference.

Nevertheless, ethnicity-specific cutoffs for waist circumference for adults have been developed by several investigators (Table 11.7); most have been optimized for the identification of adults with elevated cardiovascular risk, except those for Japanese adults, in whom a visceral adipose tissue volume > 100cm^3 was applied (Hiuge-Shimizu et al., 2012).

Table 11.7 Waist circumference (cm) for adults above which cardiometabolic risk is elevated. Japanese waist circumference values are thresholds above which visceral adipose tissue volume is > 100cm^3. The original data sources, along with this summary are given in Ross et al. (2020).

Ethnic Group	Men	Women
Japanese	≥ 85	≥90
Jordanian	≥ 98	≥ 96
Chinese	≥ 80	≥ 80
Korean	≥ 90	≥ 85
Tuisian	≥ 85	≥ 85
Iranian	≥ 89	≥ 91
Asian Indian	≥ 90	≥ 80

Most of the values in this table were derived from cross-sectional data rather than prospective studies using representative populations and were not considered in association with BMI. Of note is the wide range in high-risk waist circumference values for both adult men (80-98cm) and women (80-96cm).

In the future Ross et al. (2020) recommend conducting prospective studies using representative populations to address the

need for BMI category-specific waist circumference cutoffs across different ages, and by sex and ethnicity. Such data have been developed only for Caucasian adults by Ardern et al. (2004) and are summarized in Table 11.8. These investigators reported that in both sexes, the use of BMI category-specific waist circumference cutoffs improved the identification of individuals at high risk of future coronary events. These results were confirmed in a later study in which the prognostic performance of the Ardern waist circumference values was compared with the traditional U.S. waist circumference cutoffs associated with high cardiometabolic risk (i.e., > 88cm for Caucasian women; > 102cm for Caucasian men). Again, stratification of waist circumference cutoffs by BMI substantially improved predictions of mortality compared with the traditional waist circumference cutoffs for U.S. Caucasian adults of both sexes (Bajaj et al., 2009).

Table 11.8 Waist circumference thresholds (cm) stratified by BMI for white individuals. Subjects with measurements higher than these values have a high risk of future coronary events (based on 10-year risk of coronary events or the presence of diabetes mellitus). Data from Ross et al. (2020).

BMI category (kg/m^2)	Women	Men
Normal weight (18.5-24.9)	≥ 80	≥ 90
Overweight (25-29.9)	≥ 90	≥ 100
Obese I (30-34.9)	≥ 105	≥ 110
Obese II and III (≥35)	≥ 115	≥ 125

Waist circumference is also a highly sensitive and specific marker of accumulation of central obesity in children. Several country-specific waist circumference percentile cutoffs for children have been developed(Goran & Gower, 1999; Nagy et al., 2014; Eisenmann, 2005; Serrano et al., 2021).

Recently, international age- and sex-specific waist circumference cutoffs to define central obesity for children and adolescents aged 6-18y have also been developed (Xi et al., 2020). Based on data from 8 countries (Bulgaria, China, Iran, Korea, Malaysia, Poland, Seychelles, Switzerland), the chosen cutoff is the 90th waist circumference percentile in children with normal body weight (based on BMI). This cutoff performed well to predict cardiovascular risk when based on available data from 3 countries (China, Iran, Korea) on the presence of three or more of six cardiovascular risk factors: systolic blood pressure, diastolic blood pressure, total cholesterol, triglycerides, high-density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), and fasting glucose.

The calculated 90th percentile waist circumference values for children aged 6-18y with normal weight (i.e., excluding those who were underweight, overweight, or obese) and based on the pooled data from 113,453 children in 8 countries, are shown in the sex-specific columns (Table 11.9). However, more research is needed to further evaluate the performance of the proposed age- and sex-specific 90th percentile WC values in other populations. See Xi et al. (2020) for more details.

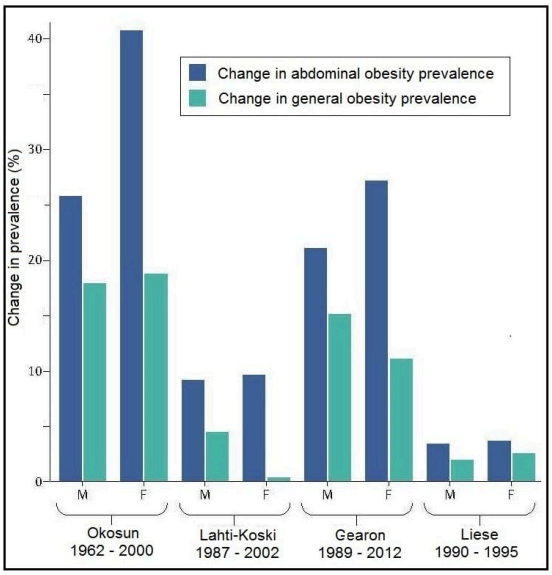


Figure 11.13 The prevalence of abdominal obesity and obesity measured in different studies. Changes in the prevalence of abdominal obesity (measured using WC) and general obesity (measured using BMI) measured in different studies during the time period indicated on the x axis. General obesity was defined as BMI ≥ 30kg/m^2. Abdominal obesity was defied as WC ≥ 88cm and ≥ 102cm for women and men, respectively. Years given (for example, 1962-2000) indicate the years in which the data were colllected. F = female; M = male. Redrawn from Ross et al. (2020).

Finally, emerging evidence suggests the relative increases in waist circumference in adults are larger than the relative increases in BMI across populations (Visscher et al., 2015). This trend appears to be independent of age, and sex and ethnicity as shown in Figure 11.13 (Ross et al., 2020), and emphasizes that a single focus on BMI > 25 or > 30kg/m^2 is likely to mask a real increase in the obesity epidemic. Clearly, waist circumference should be included along with BMI in all obesity surveillance studies in the future to ensure the phenotype of obesity that conveys the greatest health risk (i.e., abdominal obesity) is identified. This recommendation was made by the International Atherosclerosis Society (IAs) and the International Chair on Cardiometabolic Risk (ICCR) working group on visceral obesity. In addition, the working group have emphasized the importance of research to refine the WC cutoffs for a given BMI category (Table 11.8) to optimize obesity risk stratification across age, sex, and ethnicity (Ross et al., 2020).

This page titled 10.2: Assessment of body fat (11.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

10.3: Assessment of fat-free mass (11.2)

The fat-free mass consists of the skeletal muscle, non-skeletal muscle, organs, connective tissue, total body water, and the skeleton (Earthman, 2015). Based on the two-component model, once total body fat has been estimated, then fat-free mass can be determined, as shown below:

$$\text{Total body fat (kg)} = \text{body weight (kg)} \times \% \text{ body fat } /100$$
$$\text{Fat-free mass} = \text{body weight (kg)} - \text{body fat (kg)}$$

Therefore, the limitations outlined earlier when using anthropometric variables to assess percentage body fat based on the two-component model must also be considered when assessing fat-free mass. Recognition of these limitations is especially important when assessing fat-free mass in undernourished children with edema (Wells & Fewtrell, 2006; Girma et al., 2016), and obese subjects (Gutiérrez-Marín et al., 2021). In these circumstances, the assumptions used to convert from raw measurements to final body composition values (see Section 11.1.5) are often violated.

Muscle is a major component of the fat-free mass and the primary site for glucose uptake and storage as well as a reservoir of amino acids stored as protein, as noted earlier. Assessment of muscle mass can therefore provide an index of the protein reserves of the body, which can be metabolized during periods of negative nitrogen balance. Loss of muscle mass is an important criterion for the definition of malnutrition (Cederholm et al., 2019), and for the diagnosis of sarcopenia in the elderly. Depleted muscle mass in malnourished children increases risk of mortality during infections (Briend et al., 2015), whereas in older adults, sarcopenia may be associated with several negative outcomes, including falls, fractures, and mobility disorders, cognitive impairments, and mortality (Cruz-Jentoft et al., 2019).

Several simple, non-invasive anthropometric measurements based on mid-upper arm circumference (MUAC), either alone (Hu et al., 2021), or in combination with triceps skinfold thickness (i.e., arm-muscle circumference and arm-muscle area), are used as surrogates for muscle mass in both clinical and community settings. All three of these measurements have been shown to correlate with muscle mass assessed by *in vivo* laboratory-based methods such as bioelectrical impedance analysis (BIA) (Hu et al., 2021), DXA, or computed axial tomography (Heymsfield et al., 1982; Carnevale et al., 2018). As a result, they have also been used to predict changes in protein status in resource-poor settings, provided the changes are not small.

More recently, calf circumference and hand grip strength have also been recommended as tools to assess the amount and strength of muscle mass and identify older people at risk for sarcopenia in clinical practice (Chen et al., 2020). These anthropometric measurements and indices derived from them are discussed below.

11.2.1 Mid-upper-arm circumference

The arm contains both subcutaneous fat and muscle; a decrease in MUAC may therefore reflect a reduction in either muscle mass or subcutaneous tissue (or both). In some low-income countries, where the amount of subcutaneous fat is often small, changes in MUAC tend to parallel changes in muscle mass and, hence, are sometimes used as in indicators of severe and moderately severe undernutrition in young children (age < 5y) from resource-poor settings. Even in high-income settings, measurement of MUAC is recommended as one of the indicators of pediatric malnutrition by the Academy of Nutrition and Dietetics (AND) and the American Society for Parenteral and Enteral Nutrition (ASPEN) (Becker et al., 2014). MUAC measurements are particularly important for children whose weight may be affected by edema, ascites, or steroids in the lower extremities because of fluid retention (Mehta et al., 2013). Changes in the MUAC measurements can also be used to monitor progress during nutritional therapy.

The measurement of MUAC requires a minimal amount of time and equipment, and changes in MUAC are easy to detect. Therefore, increasingly, MUAC is used in emergencies such as famines and refugee crises for screening children with severe acute malnutrition (SAM) or moderate acute malnutrition. In such situations, the measurement of weight or height may not be feasible, and ages of the children are often uncertain (de Onis et al., 1997). In addition, use of a weight-based nutritional assessment (e.g., WLZ) can be misleading in children with SAM who frequently have diarrheal disease accompanied by dehydration, which lowers the weight of a child. Modi et al. (2015) showed that MUAC outperformed weight-for-length Z-score < –3 to identify SAM in children aged 6-60mo with diarrhea.

The use of fixed cutoffs to distinguish normal and malnourished children assumes that MUAC is relatively independent of age for

these children. However, the age independence of MUAC has been questioned (Hall et al., 1993; Bern & Nathanail, 1995; WHO, 1995). This has led to the use of MUAC Z-scores, that adjust for age and sex differences (Houssain et al., 2017).

Research in Somalia suggested improved concordance in prevalence estimates for acute malnutrition using MUAC-for-age Z-scores rather than MUAC alone (Custodio et al., 2018). However, Leidman et al. (2019) reported that the convergence with weight-for-height Z-score data when MUAC-for-age Z-score replaced MUAC alone was limited, based on data from population surveys from 41 countries. They concluded that the additional estimation of age, required when using MUAC-for-age Z-score, especially in humanitarian settings, was not justified. They urged the need for further research on morbidity and mortality of children with low MUAC-for-age Z-scores.

Measurement of MUAC alone has also been used to detect overweight and obesity in children and adolescents in both low- and high-income settings due to the strong relation with body weight. Craig et al. (2014) concluded that MUAC may have potential for clinical and surveillance application as an accurate indicator of overweight and fatness in children and adolescence in place of BMI that requires measurements of both weight and stature. In their study of black South African children age 5-14y, overweight was defined on the basis of BMI-for-age and overfatness from body fatness via bioelectrical impedance (BIA) estimates of body fatness. Talma et al. (2019) measured MUAC and BMI in children aged 2-18y from the fifth Dutch Nationwide Growth Study. They also concluded that in studies when weight and stature measurements are impossible, MUAC can be used as an alternative and valid measure for detecting overweight and obesity. Certainly, the results of cross-sectional data from 12 countries representing five major geographic regions of the world suggest that MUAC may be a promising tool for obesity in resource-poor settings (Chaput et al., 2017).

A major application of MUAC in older adults is as a surrogate for appendicular skeletal muscle mass, and the subsequent detection of sarcopenia (Pinheiro et al., 2020; Hu et al., 2021). MUAC is recommended because the measurement is less affected by fluid retention compared to the lower extremities, a condition that often occurs in older adults. Sarcopenia is characterized by decreases in muscle mass, as well as strength and function, all of which have multiple adverse health consequences, as noted earlier. Several studies have demonstrated that low MUAC is associated with an increased risk of all-cause mortality in adults (Weng et al., 2018). Some investigators have also used MUAC as a proxy for BMI measurements to classify adults as thin (Tang et al., 2020).

Measurement of mid-upper arm circumference

Measurements of MUAC should be made using a flexible, non-stretch tape made of fiberglass or steel; alternatively, a fiberglass insertion tape can be used. The subject should stand erect and sideways to the measurer, with the head in the Frankfurt plane, arms relaxed, and legs apart. If the subject is wearing a sleeved garment, it should be removed or the sleeves should be rolled up. The measurement is taken at the midpoint of the upper arm, between the acromion process and the tip of the olecranon (Figure 11.14).

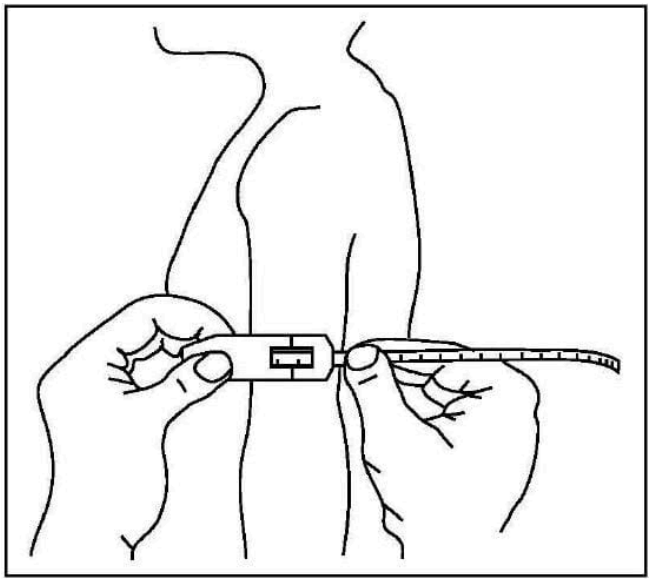


Figure 11.14 Use of insertion tape to measure mid-upper arm-circumference. Redrawn from Robbins & Trowbridge (1984).

After locating the midpoint, the arm is extended so that it is hanging loosely by the side, with the palm facing inward. The tape is then wrapped gently but firmly around the arm at the midpoint (Figure 11.14), care being taken to ensure that the arm is not squeezed. MUAC is often measured at the same time as triceps skinfold. In the WHO Multicentre Growth Reference Study, MUAC (and skinfold) measurements were taken on the left side of the body (de Onis et al., 2004).

The MUAC measurement tapes are cheap, widely available, and easy to use so that cases of SAM can be readily identified at the community level. If necessary, MUAC can be measured with subjects in the recumbent position. In this case, a sandbag is placed under the elbow to raise the arm slightly off the surface of the bed (Chumlea et al.,1984).Measurements are taken to the nearest mm.

Precision of MUAC measurements, both within and between examiners, can be high, even if the subjects are obese, provided that trained examiners and standardized methods are used. High precision is critical as MUAC varies little at any given age, with measurements tending to form a narrow symmetrical distribution, so that even small errors are significant. In the WHO Multicentre Growth Reference Study, the maximum allowable difference for the duplicate measurements of arm circumference was 5.0mm (de Onis et al., 2004).

Interpretive criteria

To classify acute malnutrition in children (6-59mos), a single MUAC cutoff irrespective of age (i.e., 125mm) is often used, as a proxy for low weight-for-height (i.e., WHZ < –2; wasting) (Leidman et al., 2019).

WHO (2009) recommends a MUAC cutoff of < 115mm to diagnose children aged 6-60mo with severe acute malnutrition (SAM), together with the presence of bilateral pitting edema, where possible. Following treatment, the WHO discharge criteria recommended for children with SAM are a MUAC cutoff of > 125mm and no edema for at least 2wks (WHO, 2009).

The U.S Academy of Nutrition and Dietetics (AND) and the American Society for Parenteral and Enteral Nutrition (ASPEN) also recommends MUAC cutoffs to classify bedbound children aged 6-60mos with undernutrition when measurements of weight and length or height are not feasible. For children classified as severely malnourished, a MUAC cutoff < 115mm is recommended, for moderately malnourished children, a MUAC cutoff of 115-124mm, and for children at risk of malnutrition, a MUAC cutoff of 125-134mm (Becker et al., 2014).

Chaput et al. (2017) suggest a MUAC cutoff of about 25cm (250mm) for both boys and girls to identify obesity in children 9-11y, based on their 12 country study data. However, from country-specific analyses, the cutoff value to identify obesity ranged from 23.2cm (boys in South Africa) to 26.2cm (girls in the UK; see Chaput et al. (2017) for more details.

There is no consensus on an optimal MUAC cutoff for thinness among adults, with reported cutoff values ranging from 17.0cm to 25.1cm. This discrepancy is in part due to the use of different BMI cutoffs to define thinness (Philpott et al., 2021). Based on a meta-analyses stratified by gender, disease states, and geographies, a cutoff of < 24.0cm (240mm) has been claimed to adequately classify thinness across adult population groups(Tang et al., 2020).

Likewise, there is no consensus on the MUAC cutoff to predict low muscle mass and diagnose sarcopenia in adults; cutoffs vary with sex, age, and possibly race-ethnicity. For community-dwelling Chinese adults > 50y, Hu et al. (2021) recommend MUAC cutoffs of < 28.6cm for men and < 27.5cm for women for predicting low muscle mass, and < 27cm for both sexes to identify sarcopenia. This cutoff was developed based on low muscle mass diagnosed using the European Working Group on Sarcopenia in Older People 2 (EWGSOP2) criteria. Whether these cutoffs are applicable for other race-ethnic groups warrants investigation.

Concern that use of a fixed MUAC cutoff would over diagnose wasting among younger children and under-diagnose among older children because of the dependence of MUAC on age, as noted earlier, led WHO to recommend the use of MUAC Z-scores, which adjust for age and sex differences. Consequently, WHO has developed MUAC-for-age reference data (Z-scores and percentiles by sex) for children aged 3-60mo for international use. The curves show both age-specific and sex-specific differences for boys and girls aged < 24mos. Numerical Z-score tables and charts for boys are also available (WHO ICGS Arm Circ.).

Reference ranges for MUAC percentiles are also available for US children and adolescents aged 1-20y based on the same population used in the CDC body mass growth charts (Addo et al., 2017). The pecentiles for females are shown in Figure 11.15. The authors also provide percentiles for males and the necessary LMS coefficients to calculate Z-scores.

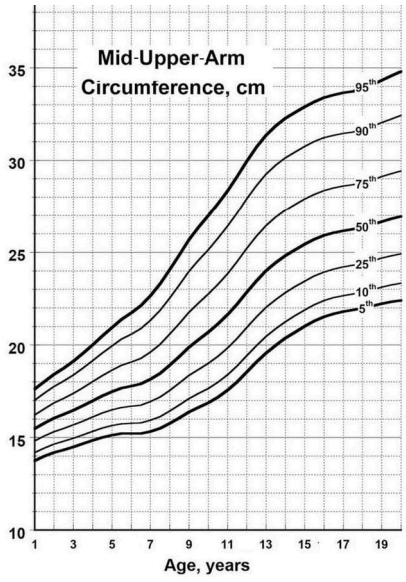


Figure 11.15 Mid-upper-arm-circumference-for-age percentiles for female US children and adolescents aged 1-20y. Redrawn from Addo et al. (2017).

It is noteworthy that for ages 2-5y, median values for the MUAC curves for the US children were comparable to those of the WHO Multicentre Child Growth Study. Predictive equations for height-for-age adjusted MUAC Z-scores for males and females are also available in Addo et al. (2017).

11.2.2 Mid-upper-arm-muscle circumference

The muscle circumference of the mid-upper arm is derived from measurements of both the MUAC and triceps skinfold thickness and is the calculated circumference of the inner circle of muscle surrounding a small central core of bone (Gurney & Jelliffe, 1973). Traditionally, in resource-poor settings, mid-upper-arm-muscle circumference (MUAMC) was used as a proxy for total body muscle mass and used to diagnose undernutrition in community surveys (Jelliffe, 1966).

Strong correlations between calculated values for MUAMC and fat-free mass estimates based on reference *in vivo* methods such DXA(Carnevale et al., 2018) and computer tomography (Lambell et al., 2021)has led to the use of MUAMC as a proxy for muscle mass in the elderly(Akin et al., 2015; Landi et al., 2010). For example, in a prospective study of older men (60-79y), MUAMC was significantly and inversely related to mortality, with the prediction of mortality being greater when MUAMC was combined with waist circumference (Wannamethee et al., 2007). In addition, in persons 80y or older, MUAMC was positively related to functional performance as well as survival. In this study, functional performance was assessed using the physical performance battery score based on three timed tests: 4-m walking speed test, the balance test, and the chair stand test (Landi et al., 2010).

Low MUAMC has also been shown to be associated with longer hospital stays in hospitalized adults. In a study by Pinto et al. (2021), such an association was independently related with undernutrition. MUAMC has also been used to detect low muscle mass in clinical and primary care settings where assessment of muscle mass using more direct *in vivo*methods such as computer tomography is not feasible (Gort-van Dijk et al., 2021).

Nevertheless, it is important to realize that the mid-upper-arm-muscle circumference is a one-dimensional measurement, whereas mid-upper-arm-muscle area is two-dimensional, and mid-upper-arm-muscle volume is three dimensional. Consequently, if the volume of the mid-upper-arm muscle declines during undernutrition or enlarges following a program of nutritional support, the mid-upper-arm-muscle circumference change will be proportionally smaller than the change in the mid-upper-arm muscle area (Heymsfield et al., 1982). Hence, MUAMC is insensitive to small changes of muscle mass that might occur, for example, during a brief illness.

Calculation of mid-upper-arm-muscle circumference

The equation for the calculation of mid-upper-arm-muscle circumference (MUAMC) is based on the same assumptions as those described for mid-upper-arm fat area (Section 11.1.4).

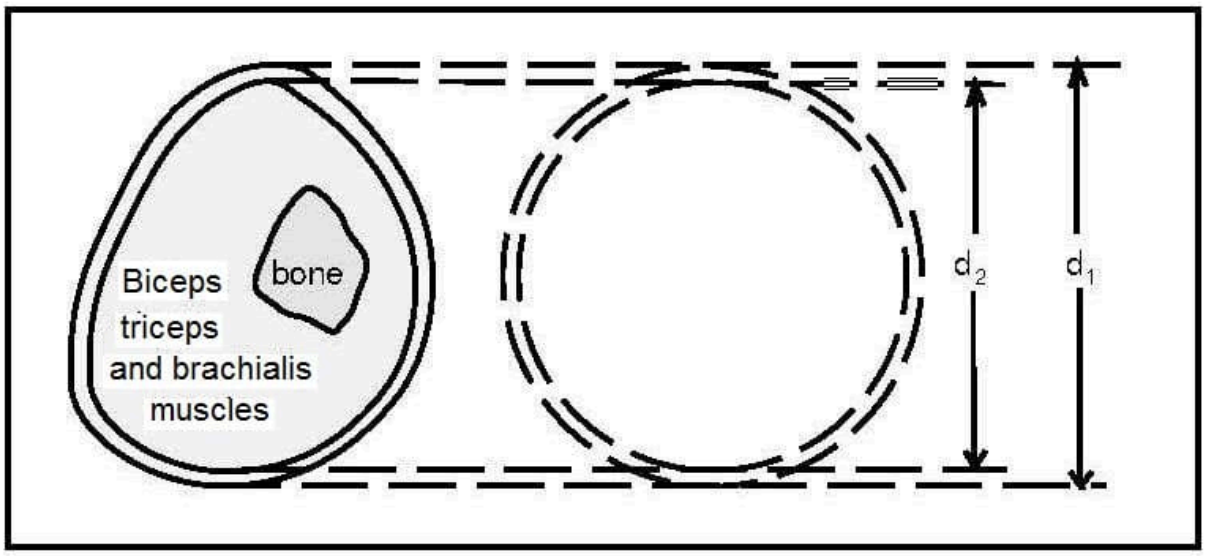


Figure 11.16 Mid-upper-arm-muscle circumference.

If MUAC = mid-upper-arm circumference, TSK = triceps skinfold, d_1 = arm diameter,

and d_2 = muscle diameter. Then:
TSK = 2 × subcutaneous fat ($d_1 - d_2$) and MUAC = π × d_1.
MUAMC = π × d_2 = π × [$d_1 - (d_1 - d_2)$]
= π × d_1 – π × ($d_1 - d_2$). Hence
MUAMC = MUAC – (π × TSK)
Note that this equation requires all measurements to be in the same units (preferably mm).

As variations in skinfold compressibility are ignored, and as the triceps skinfold of females is generally more compressible than that of males, MUAMC in females may be underestimated (Clegg & Kent, 1967). As a further complication, the MUAMC equation does not account for between subject variation in the diameter of the humerus relative to MUAC (Frisancho, 1981).

Interpretive criteria

There are no reference ranges for MUAMC for children based on the WHO Multicentre Child Growth Study or for the U.S. children used to compile the CDC 2000 BMI growth charts (Kuczmarski et al., 2000a). Some population-specific reference data based on calculated MUAMC are available for Argentinian children aged 4-14y (Oyhenart et al., 2019a).

Kuczmarski et al. (2000b) compiled MUAMC reference data for adults from the U.S. NHANES III survey (1988-1994), but only for those adults > 50y. Mean (SE) and selected percentile values of males and females for four age groups are available. During this time, values for MUAMC increased up to age 65y in women and up to middle age in men and then steadily decreased. However, secular-related changes in MUAC and triceps skinfold thickness have been reported in U.S. males and females, so caution must be used when comparing more recent MUAMC data with these earlier MUAMC reference data for U.S. adults (Frisancho, 1990).

11.2.3 Mid-upper-arm-muscle area

Mid-upper-arm-muscle area (AMA) is said to be preferable to mid-upper-arm-muscle circumference as an index of total body muscle mass because it more adequately reflects the true magnitude of muscle tissue changes (Frisancho, 1981). Several studies have examined the validity of mid-upper-arm-muscle area by comparison with *in vivo* body composition reference methods. Magnetic resonance imaging (MRI) and computer tomography (CT) are considered the gold standard methods for *in vivo* assessment of muscle mass, although use of bioelectrical impedance (BIA) and DXA is increasing in clinical settings. Unfortunately, the validity of the calculated mid-upper-arm-muscle area (AMA) as a proxy for actual arm-muscle mass is dependent on the characteristics of the study population and the *in vivo* reference method used. For example, the traditional equation appears to overestimate AMA in obese patients and may not be appropriate for undernourished children (Heymsfield et al., 1982; Rolland-Cachera et al., 1997).

Despite these limitations, calculated AMA as a proxy for arm-muscle mass has been used by several investigators. Pinto et al. (2021) reported that AMA, like MUAMC, was linked with length of hospital stay. In Caucasian adult patients with AMA values lower than the 5th percentile (i.e., indicative of depletion), the probability of being discharged from the hospital was lower. However, this finding has not been consistent in all studies (Luis et al., 2013). AMA has also served as an early indicator of deteriorating nutritional status in a longitudinal study of pediatric patients with cystic fibrosis(Ellemunter et al., 2021).

AMA usually increases in adults up to age 65y in women and up to middle age in men, and then steadily decreases, as observed for MUAMC. Caution must be used, however, when interpreting estimates of AMA among adults who are obese or those with triceps skinfold thickness that exceeds the 85th age-and sex-specific percentiles. In such persons, estimates of AMA by anthropometry are said to overestimate AMA determined by computerized tomography, the degree of overestimation varying directly with the degree of adiposity (Forbes et al., 1988).

Calculation of mid-upper-arm-muscle area

The following equation may be used to estimate mid-upper-arm-muscle area (AMA):

$$\text{Arm muscle area } = (\text{ MUAC } - (\pi \times \text{ TSK }))^2/4\pi$$

where MUAC = mid-upper-arm circumference and TSK = triceps skinfold thickness (Frisancho, 1981). Consistent units, preferably mm, should be used throughout. The index is based on the following assumptions:

- The mid-upper-arm cross-section is circular.
- The triceps skinfold is twice the average adipose tissue rim diameter at the middle of the upper arm.
- The mid-upper-arm-muscle component is circular in cross-section.
- Bone atrophies in proportion to muscle wastage during protein-energy malnutrition.

The cross-sectional areas of neurovascular tissue and the humerus are relatively small and ignored. The first three assumptions are those used in the calculations of mid-upper-arm-fat area and mid-upper-arm muscle circumference. Several investigators have reported the equation shown above overestimates arm-muscle area compared with values based on *in vivo* reference methods such as computer axial tomography (Heymsfield et al., 1982) or magnetic resonance imaging (Rolland-Cachera et al., 1997). As a result, revised equations have been developed. However, their use has been limited, in part because the corrected formulae have only been validated in selected population groups, and population-specific corrected equations may be needed.

Interpretive criteria

Age and sex-specific smoothed percentiles for AMA based on the same population of healthy U.S. children aged 1-20y used to construct the CDC 2000 BMI charts (Kuczmarski et al., 2000a) and the 2010 skinfold thickness percentiles of Addo and Himes (2010) are shown in Figure 11.17 (Addo et al., 2017). The authors also provide the necessary LMS coefficients to calculate Z-scores; see Addo et al. (2017) for more details.

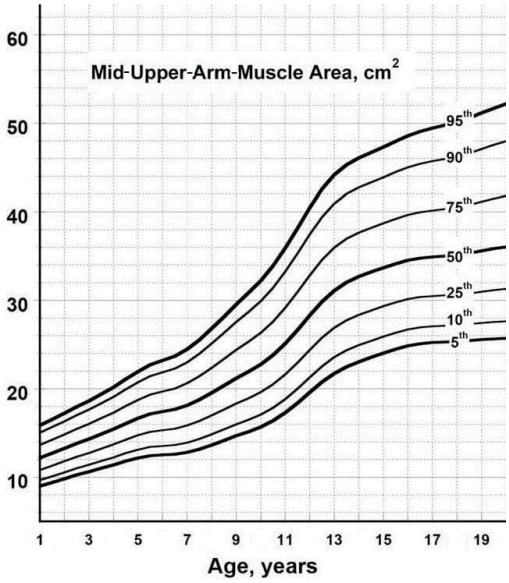


Figure 11.17 Mid-upper-arm-muscle area-for-age percentiles for female US children and adolescents aged 1-20y. Redrawn from Addo et al. (2017).

Note the percentile curves in Figure 11.17 represent the entire race-ethnicity groups that were represented in the survey, so that interpretation based on these percentile curves should take into account any potential effects of specific race-ethnicity.

Some local region-specific percentiles curves for AMA are available, including those for boys and girls aged 4-14y in Argentina (Oyhenart et al., 2019a). The mean AMA values for 3rd, 50th, and 97th percentiles are lower for the Agentinean children than their U.S. counterparts (Oyhenart et al., 2019b).

Prediction equations adjusted for the age-dependent effects of height on AMA have also been developed by Addo et al. (2017). These prediction equations may be especially helpful in populations with a high prevalence of stunted children or those in which even well-nourished children are shorter than the average height.

Reference values for AMA for U.S. adults (18-74y) or an older subset aged > 50y have not been compiled from the U.S. NHANES

III data. Earlier data on AMA percentiles (5th through 95th) are available based on the merged NHANES I and II data (Frisancho, 1990). These data are presented by age, sex, and race-ethnicity for persons 1-74y; by height for boys and girls 2-17y; and by age, sex, and frame size for adults aged 18-74y. However, secular changes in AMA in U.S. males and females have been reported, so caution must be used when interpreting data based on these older reference data.

11.2.4 Calf circumference

The quantity of skeletal muscle (i.e., muscle mass) changes over the life course: increasing during growth in childhood, stabilizing in midlife, after which muscle mass declines with ageing. In midlife in adults about 30% of the skeletal muscle is in the lower limbs, although after age 50y, the amount declines, with losses of ~1-2% per year (Cruz-Jentoft et al., 2019).

Calf circumference is often used as a surrogate marker of skeletal muscle mass. The measurement is simple and practical and was reported as the most used measurement in clinical practice to assess muscle mass, based on an international survey in 55 countries (Bruyere et al.,2016). Calf circumference has the added advantage of having a lower fat mass compared to other body sites. As a result, the impact of fat mass on the measurements will be less (Bahat, 2021).

Calf circumference is used to diagnose sarcopenia in the elderly, a condition characterized by both low muscle mass and low muscle strength, which is associated with poor physical performance (Cruz-Jentoft et al., 2019). In older adults, sarcopenia is associated with falls, fractures and mobility disorders, cognitive impairments, and mortality, as noted earlier (Cruz-Jentoft et al., 2010; 2019).

Several mechanisms may be involved in the onset and progression of sarcopenia, including protein synthesis, proteolysis, neuromuscular integrity and muscle fat content; for further details, see Cruz-Jentoft et al. (2019). However, other causes of sarcopenia besides ageing have been identified, including systemic disease, especially those invoking inflammation (e.g., cachexia), and malnutrition. For example, the Global Leadership Initiative on Malnutrition has included low muscle mass as an important criterion for the definition of malnutrition (Cederholm et al., 2019).

The use of calf circumference as a proxy for muscle mass to identify older adults with or at risk for sarcopenia is supported by both the Asian Working Group for Sarcopenia (AWGS) (Chen et al., 2020), and the European Working Group on Sarcopenia in Older People (EWGSOP) (Cruz-Jentoft et al., 2019), especially in settings where no other muscle mass diagnostic methods are available. Several investigators have compared the performance of calf circumference as an indirect anthropometric marker of appendicular skeletal muscle mass against *in vivo* reference methods. Most of these studies have been on the elderly and have reported moderate to good correlations of calf circumference against direct measurements of skeletal muscle mass using reference methods such as dual-energy X-ray absorptiometry (DXA) (Kawakami et al., 2015) and bioelectrical impedance(Gonzalez-Correa et al., 2020).

More recently, the use of calf circumference as an adequate predictor of skeletal muscle mass over the entire adult lifespan has been investigated (Santos et al., 2019). In this U.S. study of adults from the 1999-2006 NHANES survey, calf circumference was verified as a satisfactory predictor of appendicular skeletal muscle mass based on DXA (Santos et al., 2019). The impact of age, sex, and ethnicity was also examined, given the earlier reports of their impact on skeletal muscle measurements (Rush et al., 2009; He et al., 2003). The U.S. adult age groups studied were: 18-20y; 20-39y;40-59y; >60y and their ethnicities were: Mexican American; African America, White, Other. A prediction equation was developed based on calf circumference and age, sex, and ethnicity or self-reported race.

In a later study employing the same NHANES 1999-2006 dataset, calf circumference cutoff values for U.S. adults according to sex, ethnicity and race were defined as a marker to identify low muscle mass, validated by DXA measurements (Gonzalez et al., 2021). With these cutoff values, sarcopenia could be diagnosed, not only in older participants, but across the full adult lifespan. Factors confounding the calf circumference measurements across the entire adult lifespan were also identified. In addition to sex and ethnic or self-reported race identified earlier as confounders, BMI was also shown to be a very important confounder in this later study. Calf circumference values were reported to be lower for those with BMI < 18.5(kg/m^2), but higher for those overweight or obese compared to those with a normal BMI, irrespective of age, ethnicity, or race. Further, the confounding effect of adiposity could be removed by applying BMI adjustment factors for calf circumference for those participants outside the normal-weight BMI range (i.e., BMI 18-24.9) (see Table 11.11). Hence, by applying these BMI adjustment factors, the confounding effects of adi-

posity can be removed so that low calf circumference under any BMI can be identified.

Measurement of calf circumference

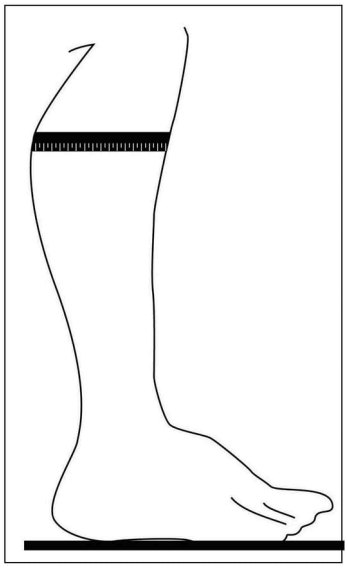


Figure 11.18 Measuring tape position for maximal calf circumference. From CDC Manual (Revised Dec. 2000)

The measurements are taken using a steel measuring tape while the subject is in a seated position. The tape is placed around the calf and moved up and down to locate the maximum circumference as shown in Figure 11.18. The tape must be held snugly but not tight and the measurement taken to the nearest 0.1cm. In the U.S. NHANES, measurements are performed on the right calf (CDC Manual, Revised Dec. 2000).

Interpretive criteria

Cutoff points for calf circumference representing low muscle mass were first developed from studies in older persons. The cutoffs recommended by the Asian Working Group for Sarcopenia (AWGS) are < 34cm for males and < 33cm for females (Chen et al., 2020).

Calf circumference cutoff values for U.S. adults ≥ 18y with a normal BMI (18.5-24.9kg/m^2) and defined by Gonzalez-Arellanes et al. (2021) are presented in Table 11.10. The cutoff values were determined by using 1 or 2 SDs below the mean (from a reference young population aged 18-39y and of normal weight) for moderately low or severely low calf circumference values, respectively. The values based on 1 SD below the mean and indicative of a moderately low calf circumference are appropriate to detect low muscle mass in adults > 65y for sarcopenia diagnosis / screening (Table 11.10).

Table 11.10 Reference and cutoff values for calf circumference according to sex, ethnicity, and race, from participants with normal BMI[3]

[1] *Reference values defined as mean values from participants aged 18-39y.*

[2] *values defined as –1 SD below the mean values.*

[3] *BMI: 18.5-24.9 kg/m*^{2}*; for other BMI groups, use the adjusting factors for correction of calf circumference (Table 11.11). Data from González-Arellanes et al. (2021).*

	Males			Females		
	n	Reference[1] mean ± SD	Cutoff[2] –1 SD	n	Reference[1] mean ± SD	Cutoff[2] –1 SD
Total - All subjects	1639	36.3 ± 2.2	34.1	1465	35.3 ± 2.3	33.0
White Non-Hispanic	633	36.6 ± 2.2	34.4	656	35.6 ± 2.2	33.4
Black Non-Hispanic	429	36.4 ± 2.2	34.2	279	35.3 ± 2.2	33.1
Mexican American	428	34.9 ± 2.1	32.8	378	33.9 ± 2.3	31.6
Other races & ethnicities	149	36.0 ± 2.1	33.9	152	34.6 ± 2.2	32.4

Table 11.11 presents an example of the BMI adjustment factors (in cm), noted earlier, for calf circumference for males with BMIs (kg/m^2) outside the normal range based on the entire lifespan (i.e., total sample) and by ethnicity and race. Corresponding adjustment factors for females are available in González-Arellanes et al. (2021). To apply the adjustment factors, add 4cm to the calf circumference measure for BMI < 18.5) or subtract 3, 7, or 12cm from the calf circumference measure for BMI 25-29, BMI 30-39, and BMI ≥ 40, respectively. These adjustment factors are derived from linear regression for calf circumference, adjusted by age, for BMI outside the 18.5-24.9 (kg/m^2).

Application of these BMI adjustment factors to the measured calf circumference and then comparison with the cutoff values given in Table 11.10 ensures the correct identification of low calf circumference under any BMI (Table 11.11).

Table 11.11 Adjustment factors for calf circumference by BMI for males outside the 18.5-24.9 BMI range, calculated from linear regression of calf circumference, adjusted by age. Data from González-Arellanes et al.(2021), who also present comparable data for females.

BMI Group (kg/m^2)	Total All subjects	White Non-Hispanic	Black Non-Hispanic	Mexican-American	Other races & ethnicities
< 18.5	+4.3	+4.7	+4.2	+4.0	+3.4
25-29.9	–3.4	–3.4	–3.4	–3.1	–3.5
30-39.9	–6.8	–6.7	–7.2	–6.4	–6.9
≥ 40	–12.0	–11.9	–12.0	–12.1	–12.2

Without these adjustments, raw calf circumference measurements would result in an underestimate of the prevalence of low calf circumference in those with BMI ≥ 25 and an overestimate in those with a BMI < 18.5.

By applying these BMI adjustment factors, a stronger correlation of calf circumference with function may be obtained, so that in the future, BMI-adjusted calf circumference measurements should be validated for predicting sarcopenia-related health outcomes such as functional impairment, falls, and mortality. In addition, use of these BMI adjustment factors and cut-offs should be employed in studies of adults in other countries (Bahat, 2021).

11.2.5 Hand grip strength

Research suggests that a measure of muscle mass such as calf circumference should be complemented with measures of muscle strength (Lauretani et al., 2003). Studies using magnetic resonance imaging have shown that muscle strength decreases by more than 50% in in the elderly (70-82y) compared to that in young men (18-30y). Moreover, only half the decrease in muscle strength that occurs with aging is accounted for by a decrease in the volume of muscle (Morse et al., 2005).

Hand grip strength is the recommended technique for measuring muscle strength by the European Working Party on Sarcopenia in Older People (EWGSOP) (Cruz-Jentoft et al., 2019) and the Asian Working Group for Sarcopenia (AWGS) (Chen et al., 2020). It is the simplest and most inexpensive method to assess muscle strength and has been introduced by AWGS to identify "possible sarcopenia" with or without reduced physical performance in both community health care and prevention settings (Chen et al., 2020). However, more studies are needed to establish whether region-specific cutoffs for the diagnosis of sarcopenia based on hand grip strength are necessary. Hand grip strength is also used as a measure of physical fitness in children (Tremblay et al., 2010; Wick et al., 2021) and athletes (Leyk et al., 2007; Pizzigalli et al., 2017).

Over the life course, hand grip strength increases, peaking in early adult life at a strength that is maintained through to midlife, after which it declines, with the loss accelerating through old age (Roberts et al., 2011). For example,

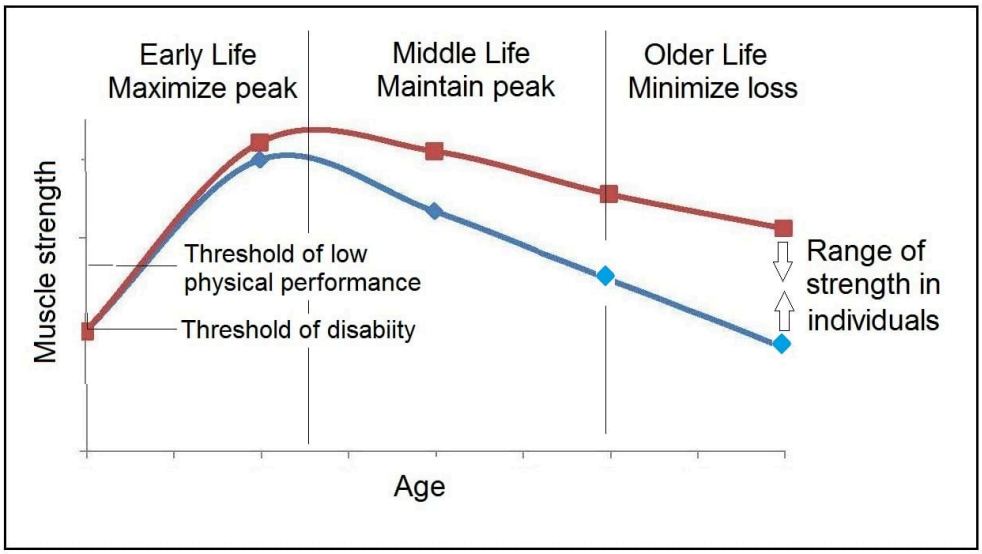


Figure 11.19. Muscle strength and the life course. Redrawn from Cruz-Jentoft et al. (2019).

in a large study of U.K. persons aged 4-90y, the prevalence of weak grip strength increased sharply with age, with a peak prevalence by age 80y of 23% in males and 27% in females (Dodds et al., 2014). In all age groups, hand grip strength is higher in males than females. Ethnic and regional variations in muscle strength may also occur, prompting studies to develop cutoffs for weak muscle strength in several geographic regions (Lera et al., 2018; Steiber, 2016; Auyeung et al., 2020).

Handgrip strength is said to be a better predictor of functional health outcomes than low muscle mass. Several functional health outcomes have been used to define optimum hand grip strength in the elderly. They include slow measured walking speed, self-reported difficulty in walking(Lauretani et al., 2003; Sallinen et al., 2011), limitations in Instrumental Activities of Daily Living, and altered physical performance. The latter may be evaluated through the Timed Up and Go test, 5-time chair stand test (> 12s), 6-meter walk (< 1.0m/s) or short physical performance battery (< 9) (Lera et al., 2018; Chen et al., 2020). Most of these relationships have been based on cross-sectional data. However, there have also been several prospective cohort studies. In some, lower handgrip strength at discharge from acute care hospitals has been associated with 30-day readmission (Allard et al., 2016), whereas in others, associations of hand-grip strength with all-cause mortality and cardiovascular mortality have been reported (Lera et al., 2018; Leong et al., 2015; Wu et al., 2017; Rantanen et al., 2003; Steiber, 2016).

Studies have also focused on the effects of modifiable behaviors such as exercise training on hand grip strength to prevent or slow adverse changes in muscle strength (Labott et al., 2019). Bohaanon (2017) has cautioned that relatively large percentage changes in grip strength may be necessary to conclude with confidence that a real change has occurred in muscle strength over time in response to treatment or an intervention. The EWSOP2 group 1 have emphasized that more studies are needed to establish whether the current recommended cutoffs improve the prediction of outcomes most sensitive to response to treatment for sarcopenia (Cruz-Jentoft et al., 2019).

Measurement of hand grip strength

Accurate measurement of hand grip strength requires use of a calibrated handheld dynamometer. The hydraulic hand dynamometer Jamar measures grip force (kgf) and is accepted as the gold standard instrument. The Jamar dynamometer is small and portable, although relatively heavy (i.e.,1.5 lb), with a dial that reads force in both kilograms and pounds, and allows assessment to the nearest 1kg or 2.5 lb. In Asia, the spring-type dynamometer (Smedley) that detects the amount of spring tension (kgf) is more widely used. Data generated by these two devices are not comparable (Kim & Shinkai, 2017). Nevertheless, the AWGS 2019 recommend using either device, provided standard measurement protocols for the device are followed (Chen et al., 2020).

Measurement of grip strength obtained by dynamometry appears to have good to excellent relative test-retest reliability, even among older adults (Bohannon, 2017). Dynamometers should be calibrated every 4-6mos to maintain longitudinal validity. The methods used to measure grip strength vary. Consequently, Roberts et al. (2011) have developed a standardized method to increase the precision of measurements within any given study, and to facilitate comparison of results across studies. Nevertheless, other factors unrelated to muscle, such as motivation or cognition, may hamper the correct assessment of muscle strength. Details of this standardized method are outlined in Box 11.1.

Box 11.1 Standardized method for grip strength via Southampton protocol based on the JAMA dynamometer (Figure 11.20).

- Sit the participant comfortably in a standard chair with legs, back support and fixed arms. Use the same chair for every measurement.
- Ask them to rest their forearms on the arms of the chair with their wrist just over the end of the arm of the chair—wrist in a neutral position, thumb facing upwards.
- Demonstrate how to use the Jamar handgrip dynamometer to show that gripping very tightly registers the best score.
- Start with the right hand.
- Position the hand so that the thumb is round one side of the handle and the four fingers are around the other side. The instrument should feel comfortable in the hand. Alter the position of the handle if necessary.
- The observer should rest the base of the dynamometer on the palm of their hand as the subject holds the dynamometer. The aim of this is to support the weight of the dynamometer (to negate the effect of gravity on peak strength), but care should be taken not to restrict its movement.
- Encourage the participant to squeeze as long and as tightly as possible or until the needle stops rising. Once the needle stops rising the participant can be instructed to stop squeezing.
- Read grip strength in kilograms from the outside dial and record the result to the nearest 1kg on the data entry form.
- Repeat measurement in the left hand.
- Do two further measurements for each hand alternating sides to give three readings in total for each side.
- The best of the six grip strength measurements is used in statistical analyses so as to encourage the subjects to get as high a score as possible.
- Also record hand dominance, i.e. right, left or ambidextrous (people who can genuinely write with both hands) and the equipment model used.

From Roberts et al. (2011).



Figure 11.20 The Jamar hydraulic hand dynamometer.

Note that in the U.S. NHANES studies the Smedley dynamometer has been used. The U.S. protocol recommends standing unless participants are unable to stand unassisted, with full elbow extension rather than sitting with 90° elbow flexion as described in Box 11.1 The AWGS (2019) also recommend standing, where possible, with full elbow extension for the Smedley dynamometer but sitting with 90° elbow flexion when the the Jamar dynamometer is used (Chen et al., 2020).

Interpretive criteria

As noted earlier, although the measurements generated from dynamometers differ depending on the device used, currently dynamometer-specific cutoff values are not recommended because of insufficient comparative data (Chen et al., 2020). Differences also exist in both the methods and functional outcomes used to define cutoff values for dynamometry.

For Asian elderly, the AWGS group recommends diagnostic hand grip cutoffs for weak muscle strength indicative of "possible sarcopenia" as < 28kg for men and < 18kg for women (Chen et al., 2020). These cutoff values are based on the lowest quintile for muscle strength based on seven community-based cohorts in East and Southeast Asia aged > 60y (Auyeung et al., 2020).

For European elderly, the EWSOP2 group have defined hand grip cutoffs for weak muscle strength indicative of sarcopenia as < 27kg for men and < 16kg for women (Cruz-Jentoft et al., 2019). These are lower than the AWGS cutoffs, and are based on the U.K. data in which weak grip strength is based on strength at least 2.5 SDs below the gender-specific mean from a healthy and young U.K. reference population (Dodds et al., 2014).

For Chilean community dwelling adults > 60y, the cutoffs proposed for weak muscle strength were < 27kg for men and < 15kg for women. These were based on the 25th percentile values (by sex) for hand grip strength for these elderly persons. Lera et al. (2018) also performed survival analysis on follow-up data of 9.2y in a subgroup of these elderly participants; hand grip values below 25th percentile were associated with an increased risk of all-cause mortality.

For Germans aged 17-90y, cutoff risks were < 33kg for men and < 21kg for women, defined as 2 SD below the sex-specific peak mean value for grip strength across the life course (Steiber, 2016). Thresholds for a critically weak grip strength associated with elevated mortality risk were defined as 1SD or more below the standardized mean hand grip strength. These cutoffs were defined using survival analysis from an 8y follow up on a restricted sample of older individuals aged 55-90y.

Several sets of region-specific normative reference values for hand grip strength are available. Examples across the life course from the U.K, U.S. Germany, and the U.S. are discussed briefly below. In the United Kingdom for example, percentiles (10th, 25th,

50th, 75th, 90th) by sex for grip strength based on cross-sectional observations in persons aged 4-90y from 12 U.K. studies have been compiled (Dodds et al., 2014). A subset of these results are presented in Table 11.12.

Table 11.12 Normative values for grip strength for UK Males. Data from Dodds et al. (2014).

		Centiles (kg)			
Age(y)	n	10th	50th	90th	Mean (SD)
10	3222	12	17	22	17.2 (4.1)
20	354	30	40	52	41.5 (7.3)
30	984	38	51	64	51.6 (9.6)
40	880	38	50	63	50.3 (10.3)
50	820	35	48	60	47.6 (10.1)
60	2683	33	45	56	44.6 (9.2)
70	3286	29	39	49	39.1 (8.1)
80	1115	23	32	42	32.2 (7.3)
90	431	16	25	33	24.7 (6.8)

In the United States, data for height (mean, SD), weight (mean, SD), mean (SD) handgrip strength (kg) and the 5th, 10th, 50th, 25th and 50thpercentiles by side of hand (dominant and non-dominant), sex, and age (3-17y) are presented(Bohannon et al., 2017). The cross-sectional data were from the US National Institutes of Health Toolbox project conducted in 2011. The standardized method for measuring hand grip strength recommended by Roberts et al. (2011) using a Jamar dynamometer, was used in this U.S. study.

Later, Wang et al. (2018) used the same U.S. dataset to compile reference values for US adults 18-85y. Again, data for height, weight, hand grip strength (kg) and percentiles (10th, 25th, 50th, 75th and 90th) by sex, side of hand (dominant and non-dominant) and 13 age groups (limited to 5-year spans, except the strata 18-24y) are presented. No cutoffs for hand grip strength based on the data for children or adults were derived.

Table 11.13 Normative reference values for handgrip strength for German women 40-44y. Abstracted from more comprehensive data for men and women aged 17-90y. Steiber(2016)

[1] Mean –1 age-group-specific SD

Height (cm)	Mean HGS (kg)	Threshhold[1] (kg)
150-154	31.5	25.3
155-159	32.7	26.4
160-164	33.7	27.4
165-169	34.8	28.6
170-174	35.8	29.6
175-179	37.1	30.8
180-184	38.0	31.8

In Germany, the normative reference values are based on a nationally representative sample of healthy participants aged 17-90y. Mean values for handgrip strength across seven height groups (within each age group: 17-19y, 20-24y, 25-29y, 30-34y, 35-39y, 40-44y, 45-49y) were calculated and presented as sex-specific reference values, stratified by age and body height. For example, the reference value for 40-44y women with a height of 165-169cm is 34.8kg; this value increases by about 1kg for every 5cm of additional height, as shown in Table 11.13 (Steiber, 2016).

In some regions, reference values of handgrip strength have been restricted to certain age groups. For example, in community-based Asian cohorts (Auyeung et al., 2020), only population data for those age > 60y were included, with the lowest quintile (kg) and mean (SD) kg by sex and age group presented, as shown in Table 11.14.

Table 11.14 Sex- and age-specific lowest quintile and means of handgrip strength (kg) in subjects age > 60y from community-based Asian cohorts. Data abstracted from Auyeung et al. (2020)

Age (y)	n	Lowest quintile (kg)	Mean (SD) (kg)
Men			
60-69.9	5319	32.7	37.9 (6.5)
70-79.9	5317	28.0	33.3 (6.3)
≥ 80	1554	23.6	28.4 (6.2)
Women			
60-69.9	6384	20.0	23.6 (4.6)
70-79.9	6009	17.8	21.1 (4.5)
≥ 80	1761	14.7	18.3 (4.5)

In Chile, percentiles by sex and age group are also restricted to community-dwelling persons > 60y (Lera et al., 2018), whereas in Canada data from the 2007-2009 Canadian Health Measures Survey included only children, with the mean and median values aged 7-10y; 11-14y; 15-19y (Tremblay et al., 2010).

Note that because all these normative reference values are based on cross-sectional data, they are likely to underestimate individual decline and hence should not be used to monitor the trajectories of individuals. Moreover, as with all cross-sectional studies, such a design limits the degree to which causal and age-related inference can be drawn (Perna et al., 2015).

This page titled 10.3: Assessment of fat-free mass (11.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

11: EVALUATION OF ANTHROPOMETRIC DATA (CHAPTER 13)

Abstract

Anthropometric indices of growth and body composition are compiled from two or more anthropometric measurements. The indices most frequently used for children include weight-for-height, height-for-age, weight-for-age, body mass index (BMI)-for-age, and mid-upper-arm circumference (MUAC)-for age. The methods recommended to evaluate these anthropometric indices use either percentiles or Z-score calculated in relation to the distribution of the anthropometric indices for a reference population. The World Health Organization (WHO) recommends for international use the prescriptive WHO Child Growth Standards for those age 0–5y and the WHO Growth Reference for those age 5–19y. New international prescriptive growth standards are also available for gestational weight gain, fetal and newborn growth, and postnatal growth for preterm infants. In clinical settings, serial anthropometric measurements are used to track an individual's growth pattern, and identify abnormal changes in growth and their response to treatment. In public health, screening systems, that include at least one anthropometric measurement or index in combination with one or more reference limits or cutoff points, can be used to identify "at risk" individuals. At the population level, screening can be used to compare and monitor the prevalence and severity of malnutrition within and across countries, and to monitor progress. Examples include use of height-for-age <–2 Z-score to identify and monitor stunting, weight-for-height <–3 Z-score or MUAC < 115mm to identify and monitor severe, acute, malnutrition in children 6–60mos, BMI-for-age Z-score or BMI for overweight and obesity in children and adults, respectively, and a composite index of anthropometric failure (CIAF). Public health applications of anthropometry discussed in this chapter include: (a) international comparisons of growth; (b) identifying determinants and consequences of malnutrition; (c) targeting interventions; (d) assessing responses to interventions; and e) nutritional surveillance. Several methods recommended for evaluating anthropometric indicators are also emphasized, the choice depending on the objectives and design of the study.

11.1: Introduction (13.0)

11.2: Anthropometric indices and their modes of expression (13.1)

11.3: Use of anthropometric indices in clinical settings (13.2)

11.4: Screening systems to identify individuals at risk in a population (13.3)

11.5: Applications of anthropometry in public health (13.4)

This page titled 11: Evaluation of Anthropometric Data (Chapter 13) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Introduction (13.0)

Standardized methods of evaluating anthropometric indices are essential to assess nutritional status and to identify malnutrition in individuals or populations. Two methods are used based on Z-scores or percentiles, both calculated in relation to the distribution of the anthropometric indices for a reference population. The World Health Organization (WHO) recommends the use of the WHO Child Growth Standards for young children0–5y internationally. They were compiled using a prescriptive approach depicting physiological human growth under optimal conditions. Hence, they represent how young children *should* grow, rather than as a "reference" describing how children *do* grow. The WHO Child Growth Standards are recommended for international use because the effects of ethnic and genetic differences on the growth of infants and young children are small compared with the environmental, nutritional, and socio-economic effects on growth, some of which may persist across generations. The reference population recommended for assessing the growth of older children is the WHO Growth Reference data for school-age children and adolescence.

A series of prescriptive standards for monitoring fetal, newborn growth, and gestational weight gain have also been developed by the INTERGROWTH-21st project for international use. This project adhered to the WHO recommendations for assessing human size and growth(WHO Working Group on Infant Growth, 1995), and followed healthy pregnant women longitudinally from 9wks of fetal life to 2y (Papageorghiou et al., 2018). Populations from urban areas of eight countries in which maternal health care and nutritional needs were met (Brazil, China, India, Italy, Kenya, Oman, the UK and the USA) were involved to ensure universal multi-ethnic growth standards were generated that represent how fetuses should grow (i.e., the standards are prescriptive). Box 13.1 lists the prescriptive growth standards generated from the INTERGROWTH-21stproject with the associated references.

st international standards for monitoring growth and development from early pregnancy to 2y.

- Fetal growth standards based on serial ultrasound measurements: Papageorghio et al. (2014).
- Estimated fetal weight standards: Stirnemann et al. (2017).
- Fetal growth velocity standards from the Fetal Growth Longitudinal Study: Ohuma et al. (2021).
- Gestational weight gain standards based on women enrolled in Fetal Growth Longitudinal study: Ismail et al. (2016).
- Newborn weight, length, and head circumference by gestational age and sex: Villar et al. (2014).
- Postnatal growth standards for preterm infants: Villar et al. (2015).

In clinical setting, serial anthropometric measurements can be used to track the pattern of growth and body composition over time. They can also be used to identify and classify an individual as "malnourished" or of "normal" nutritional status by comparison with either predetermined reference limits, often termed cutoff points. Individuals may also be classified into risk categories indicative of the severity or type of malnutrition.

However, serial measurements are often not possible in public health. Instead, screening systems, which include at least one anthropometric measurement or index in combination with the recommended reference limit or cutoff point are used to identify individuals in the population "at risk" to malnutrition. Examples of measurements or indices used in this way and discussed below include weight-for-height, body mass index (BMI), MUAC (to identify severe acute malnutrition), and a composite index of anthropometric failure (CIAF). Data derived from screening systems can be used to compare and/or monitor the prevalence of malnutrition across populations, and to assess the need for, or impact of, interventions both within and across countries. In some cases, reference limits can be combined with "prevalence thresholds" to describe and compare populations according to levels of severity of malnutrition, and identify priority populations for action. Prevalence thresholds can also be used by governments to trigger action, and to monitor and evaluate the impact of interventions.

Anthropometry can also be used to characterize the nutritional status of a population or to highlight changes in their nutritional status with successive measurements over time (i.e., surveillance). To achieve these objectives, a comparison of the frequency distribution of anthropometric indices of the entire population in relation to the recommended reference data using Z-scores (in developing countries) or percentiles is recommended.

Details of the methods recommended by WHO for evaluating anthropometric indices, followed by a review of the systems used to identify and classify risk of malnutrition at the individual and population levels, are discussed below. Examples of the applications

of anthropometry in public health are also included together with details of the designs and statistical procedures that can be used to evaluate intervention programs.

This page titled 11.1: Introduction (13.0) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Anthropometric indices and their modes of expression (13.1)

Details of the standardized procedures for the measurements of body size and body composition are outlined in Chapters 10 and 11. Raw measurements, however, have little meaning unless they are related to, for example, the age or sex of an individual (WHO, 1995). Hence, the correct interpretation of anthropometric measurements requires the use of anthropometric indices. They are usually compiled from two or more raw measurements. Details of indices of body size and body composition are also available in Chapter 10 and 11, respectively. In children, the four most commonly used anthropometric indices are weight-for-height, height-for-age, weight-for-age, and body mass index — defined as weight / (height)^2 (kg/m^2).

Weight-for-height measures body weight relative to height. Low weight-for-height is described as “thinness” and reflects a pathological process referred to as “wasting”. It arises from a failure to gain sufficient weight relative to height or from losing weight. High weight-for-height in children is referred to as “overweight” and arises from gaining excess weight relative to height or from gaining insufficient height relative to weight.

Height-for-age is a measure of achieved linear growth and can be used as an index of past nutritional or health status. Recumbent length is measured in infants and children less than 2y, and height in older children. Low height-for-age is defined as “shortness” and reflects either normal variation or a pathological process involving a failure to reach a linear growth potential. The outcome of the latter process is referred to as “stunting” or the gaining of insufficient height relative to age (WHO, 1995). The etiology of stunting is multifactorial, often resulting from a combination of extended periods of inadequate food intake, poor dietary quality, and increased morbidity.

Weight-for-age reflects body mass relative to chronological age. Low weight-for-age is described as “lightness” and reflects a pathological process referred to as “underweight”, arising from gaining insufficient weight relative to age, or losing weight. High weight-for-age can be described as “heaviness” but this term is seldom used: high weight-for-height is more useful as a proxy for “overweight” (WHO, 1995).

A major limitation of weight-for-age is that it reflects both weight-for-height and height-for-age. It fails to distinguish tall, thin children from those who are short with adequate weight. Thus, children with a low weight-for-age may be genetically short or their low weight-for-age may result from “stunting” or nutritional growth failure. Consequently, weight-for-height and height-for-age are the preferred anthropometric indices of body size for children because, in combination, they can distinguish wasting and stunting.

Table 13.1 The relative usefulness of different anthropometric indices on a scale from 1 (excellent) to 4 (poor). [a] Depends partially or et al. (1994).

	Wt. for ht.	Ht. for age
Usefulness in populations where age is unknown or uncertain	1	4
Usefulness in identifying wasted children[a]	1	4
Sensitivity to weight change over a short period of time	1	4
Usefulness in identifying stunted children	4	1

Body mass index (BMI) is the internationally recommended index of overweight and obesity in children and adults and calculated as:

$$\text{weight (kg)} / \text{ height (m)}^2$$

An increased BMI-for-age in childhood and adolescence is associated with a higher percentage of body fat, which is a known

risk factor for cardiovascular disease (de Onis and Lobstein, 2010). However, BMI alone does not distinguish between weight associated with muscle, and weight associated with body fat. Hence, in some circumstances, an elevated BMI may result from adiposity, muscularity, or edema. Furthermore, BMI is insensitive to the actual distribution of body fat (Thomas et al., 2012). For more details, see Chapter 10. Table 13.1 summarizes the information that can be obtained from each of the four growth indices discussed above.

Note that to interpret any single measurement of either weight or stature in relation to the reference data, the exact age of the child at the date of the measurement must be known. Software such as WHO AnthroPlus can calculate exact ages in decimal fractions of a year, from birth and visit dates. In the event that documentary evidence of the date of birth is not available, it may be necessary to obtain at least the month and year of birth using a local event calendar. Information on the development of a local event calendar is available from FAO (2008).

For studies of both individuals and populations, the anthropometric indices can be compared to the reference population using percentiles or Z-scores derived from the reference data. The reference populations recommended are the WHO Child Growth Standards (WHO, 2006) for infants and children 0–5y, and the WHO Growth Reference data for school-age children and adolescents 5–19y (de Onis et al., 2007).

WHO Multicentre Growth Reference Study Group (2006). WHO Child Growth Standards based on length/height, weight and age. Acta paediatrica (Oslo, Norway : 1992). Supplement, 450, 76–85.

The WHO Child Growth Standards were based on a multicenter growth reference study for infants and children aged 0–5y from six diverse geographic sites (Brazil, Ghana, India, Norway, Oman, and the U.S.), as noted earlier. The study combined a longitudinal design from 0–24mos with a cross-sectional study of children aged 18–71mo from the same sites (de Onis et al., 2004). A prescriptive approach was taken for the development of these growth standards with the goal of providing a single international reference that represents the best description of physiological growth for all children ≤ 5y and to establish the breastfed infant as the normative model for growth and development.

The WHO Growth Reference for school-age children and adolescents 5–19y was developed from the 1977 National Center for Health Statistics data set (Silveira et al., 2011), supplemented with data from the WHO Child Growth Standard. The statistical methodology used to construct this reference was the same as that used for the WHO Child Growth Standard. More details of these child growth standards are available from WHO.

In most industrialized countries, percentiles are used, whereas in low-income countries, the use of Z-scores is recommended. The proportion of individuals with indices below predefined reference limits based on percentiles or Z-scores can also be assessed and are widely used as indicators of community status.

In some circumstances, percentiles and Z-scores cannot be calculated and, instead, the anthropometric indices are expressed in terms of percent-of-median. Percent-of-median, however, provides only limited information on the relative position of the value within the population and as a consequence, is rarely used today.

13.1.1: Percentiles

A percentile refers to the position of the measurement value in relation to all the measurements for the reference population, ranked in order of magnitude. Figure 13.1 is a cumulative frequency distribution of the heights of boys aged 13y, illustrating the use of percentiles. In this figure, a height of 152cm represents the 50th percentile. This means that 50% of the boys in the reference population have a height at or below this value and 50% at or above this value. Similarly, 10% of the boys have a height at or below the 10th percentile that, in Figure 13.1, falls at 141.8cm. For data with a Gaussian distribution (e.g. height for age), the 50th percentile corresponds to both the mean and the median; for skewed data (e.g. weight-for-age), the 50th percentile corresponds to the median.

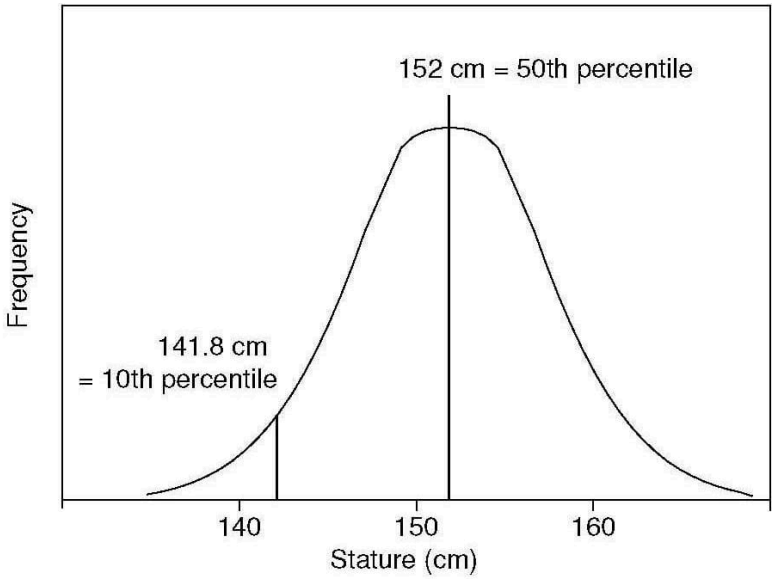


Figure 13.1. Frequency distribution for stature in 13-y-old boys.

The percentile for an individual of known age and sex can be calculated exactly, if the numerical percentile values are available for the reference data. Computer programs (e.g. WHO AnthroPlus for personal computers) are available to calculate exact percentiles (1st to 99th percentiles) based on the WHO Child Growth Standards (0–5y) (WHO, 2006) and the WHO Reference (2007) for school-aged children and adolescents (5–19y) (de Onis et al., 2007). Alternatively, the updated US Center for Disease Control and Prevention (CDC) 2000 Growth Reference in the United States (Kuczmarski et al., 2000) can be used from which numerical percentile values can be calculated based on the method described by Vidmar et al. (2004). Alternatively, the percentile range within which the measurement of an individual falls can be read from sex-specific graphs or tables of the appropriate reference data.

For children (0–5y), percentile charts are for: weight-for-length from 45–110cm; weight-for-height from 65–120cm; weight-for-age (0–60completed months); length / height-for-age (0–60mos); arm-circumference-for age (0–60mos); BMI-for-age (0–60mos); head-circumference-for-age (0–60mos); triceps skinfold-for-age (0–60mos); subscapular-skinfold-for-age (0–60mos). Growth charts are available from WHO. For field use, simplified field tables are also available from WHO. For the WHO Growth Reference 2007 for school-aged children and adolescents (5–19y), percentile charts for weight-for-age (5–10y) and height-for-age and BMI-for-age (5–19y) are discussed in de Onis et al. (2007).

Sometimes, when calculating percentiles, adjustments are made for parental stature, to distinguish between genetic and pathological effects in a child who is unusually short. The magnitude of the adjustment tends to be greater for older children. An example of adjustments based on mid-parent height is given in Himes et al. (1985).

Adjustments are also important when evaluating the growth percentiles of both low birth weight (i.e., < 2500g) and preterm infants (i.e., < 38wk gestation). Specialized growth charts are available for determining the growth percentiles of preterm infants (Villar et al., 2015).

For population studies, the number and percentage of individuals falling within specified percentiles of the reference data can be tabulated or presented graphically to provide an estimate of “relative status”, as shown in Figure 13.2.

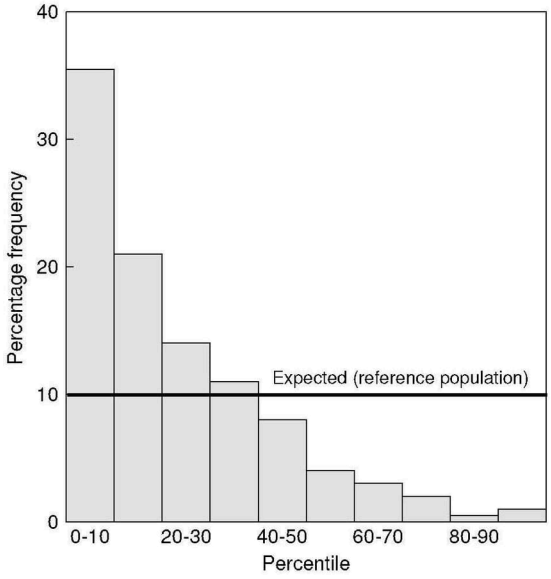


Figure 13.2. Percentage distribution of children by weight/height percentiles. From the 1974 Sahel nutrition studies based on measurements of 798 children. Redrawn from: WHO (1983).

This approach highlights critical features of the distribution of the study population compared to that of the reference population (WHO, 1986).

The percentile reference limits commonly used for designating individuals as "at risk" to malnutrition are either below the 3rd or 5th percentiles or above the 97th or 95th percentiles. The limits chosen depend on the reference data used.

Use of percentiles is recommended for the evaluation of anthropometric measurements from relatively well-nourished populations from industrialized countries, as no errors are introduced if the data have a skewed distribution. Weight-for-age, weight-for-height, and many circumferential and skinfold indices have skewed distributions.

Percentiles are not recommended for evaluating anthropometric indices of individuals or populations from low-income countries if reference data from industrialized countries, such as the US CDC 2000 (Kuczmarski et al., 2000), are used. In such circumstances, an individual, or many participants with large deficits in weight and height, may have indices below the extreme percentile of the reference distribution (i.e. below 3rdor 5th percentile), making it difficult to characterize the magnitude of the deficit for the individual or accurately classify large numbers of individuals (Waterlow et al., 1977). Moreover, because data from a population, expressed in terms of percentiles, are often not normally distributed, such populations cannot be correctly described in terms of means and standard deviations of the percentiles.

13.1.2 Z-scores

The World Health Organization recommends the use of Z-scores for evaluating anthropometric data from low-income countries (Gorstein et al., 1994), because Z-scores can be calculated accurately beyond the limits of the original reference data. This is an advantage in low-income countries because individuals with indices below the extreme percentiles of the reference data can then be classified accurately.

The method measures the deviation of the anthropometric measurement from the reference median in terms of standard deviations (SD) or Z-scores. The score is a measure of an individual's value with respect to the distribution of the reference population.

Z-scores are calculated differently for measurements that are distributed normally (e.g., height- for-age) and non-normally in the reference population (e.g., weight-for-age). The Z-score for a measurement that is normally distributed corresponds to:

$$\text{Z-score} = \frac{(\text{ observed value }) - (\text{ median reference value })}{\text{standard deviation of the reference population}}$$

As an example, a boy called Sam is 96.1cm tall, and aged 2y and 4mo

- the observed value is Sam's height (96.1cm).
- the median reference value for height of all boys who are Sam's age (2y and 4mo) is 90.4cm.
- the standard deviation of boy's height at 2y and 4mos is 3.3. The standard deviation of boy's height can be thought of as the average of differences from the median for each member of the reference population.
- inserting the above numbers in the formula, Sam's Z-score is calculated as:

$$\frac{(96.1 - 90.4)}{3.3} = 1.73$$

Note: Sam's height-for age could also be plotted manually on the WHO Child Growth Chart for boys aged 2–5y shown in Figure 13.3.

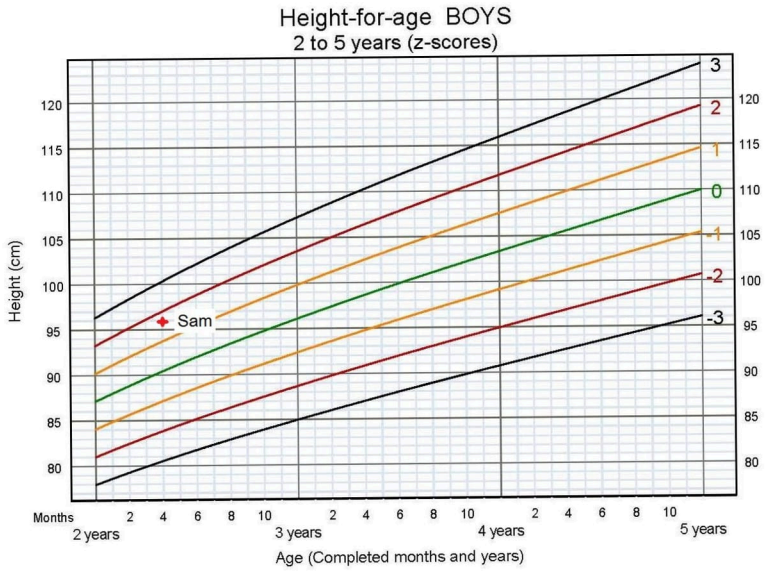


Figure 13.3. From: WHO Growth Chart for boys age 2–5y (WHO child-growth-standards)

However, the above formula cannot be used for measurements that are not normally distributed such as weight-for-age, weight-for-height, and BMI-for-age, because the distances between adjacent Z-scores are not constant. As a result, an LMS formula is used to calculate the Z-scores for measurements that are skewed. For more details, see Cole (1990)and Vidmar et al. (2004). Note that for measurements that fall beyond –3 and +3 Z-scores a modified version of the LMS formula must be applied (incorporated within WHO Anthro software).

These calculations can be carried out using the WHO software program (WHO AnthroPlus 2009) which uses the WHO Child Growth Standard for those 0–60mos (WHO, 2006) and the WHO Growth Reference for children and adolescents aged 5–19y (de Onis et al., 2007) as the reference populations. As noted earlier for percentiles, for clinical use, the Z-score range within which the measurement of an individual falls can also be read from graphs or tables of the appropriate reference data for the same indicators for infants and children age 0–60mos and children and adolescents age 5–19y. Simplified field tables are also available.

WHO has also developed a tool for the application of the WHO Child Growth Standards which includes instructions of how to take the measurements, interpret growth indicators, investigate causes of growth problems, and how to counsel caregivers. A child age calculator is available as part of the course materials for the trainers and available in WHO Regional Offices. An anthropometry training (video) is also available. Motor milestone windows of achievement have also been developed and are available from (WHO.)

Guidelines have also been developed on the use of the fixed exclusion method to identify implausible Z-score value for each anthropometric growth index based on the WHO Child Growth Standard. Z-score values outside the following intervals are considered implausible:

- HAZ (below −6, or above +6);
- WHZ (below −5, or above +5);
- WAZ (below −6, or above +5).

When using WHO AnthroPlus these "implausible" values for each anthropometric Z-score are flagged independently and the percentage of implausible values for each index calculated. Note: the date of birth (month and year) is required to calculate Z-scores and errors in this date information are a major source of implausible values. A percentage of implausible values that exceeds 1% is said to indicate poor data quality. However, inaccurate values may still exist within the rather generous flag range set by WHO, so even a low percentage does not necessarily imply adequate data quality.

WHO also recommends calculating the standard deviation (SD) of Z-scores for each anthropometric growth index as an additional assessment of data quality. This is important because an inflated SD arising from poor quality data results in prevalence estimates that are likely to be overestimated. Currently, WHO has not set cut-offs at which SDs for each anthropometric index are related to poor data quality. Instead, WHO suggests that strata specific SDs that are greater than the SD for the national estimate should be examined and highlighted in any survey report. More details on standardizing the analysis, interpretation, and reporting of anthropometric data for infants are available from (WHO).

Z-scores for anthropometric indices using alternative reference populations such as U.S. CDC 2000 Growth Reference in the United States, (Kuczmarski et al., 2000) can also be calculated based on the method described by Vidmar et al. (2004).

In population studies, the number and the proportion of individuals within a specified range of Z-scores for each age and sex group can be tabulated or presented graphically to provide an estimate of "relative" status, as shown in Figure 13.4

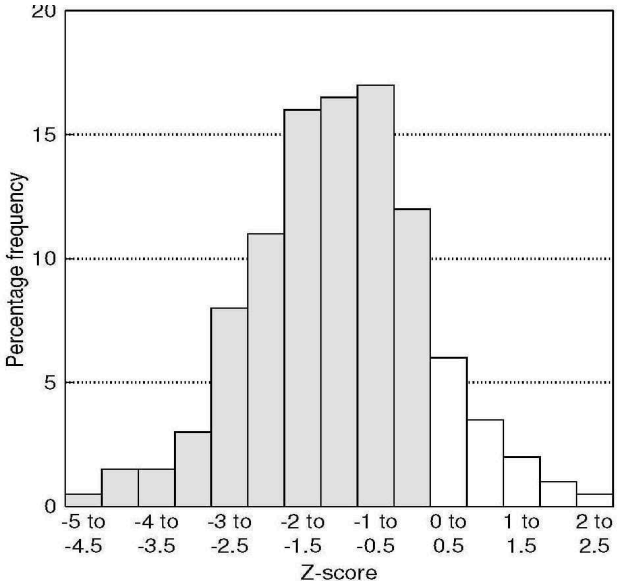


Figure 13.4. Frequency distribution of Z-scores of height-for-age for 1395 children measured during the Sahel nutrition studies and who were 12–23.9mo. Redrawn from: WHO (1983).

and as noted previously for percentiles (Figure 13.2). In addition, the proportion of individuals with anthropometric indices below or above some predetermined reference limits defined by Z-scores can also be determined. Often scores of below –2 or above +2 are used to designate individuals with either unusually low or unusually high anthropometric indices. This approach is used because statistically 95% of the international reference population fall within the central range. Theoretically the proportion of children with a Z-score less than –2 or greater than +2 in a study population should be ≈ 2.3%. Clearly, if the proportion in the study population with such low or high Z-scores is significantly greater than this, then the study population is seriously affected.

The WHO Global Database on Child Growth and Malnutrition uses a Z-score less than –2 to classify low weight-for-age as "underweight", low length/height-for-age as "stunted", and low weight-for-length as "wasted" (WHO, 2006). In children aged 0–5y, a Z-score above +2 is used to classify weight-for-height as "overweight" and as "obese" with a weight-for-height Z-score above +3.

Figure 13.5 presents the prevalence of wasting, stunting, and overweight based on these WHO indicators in children at age 1 year in 1982, 1993, 2004, and 2015 in population-based cohorts in Pelostas, Brazil,1982–2015. The prevalence of stunting declined by 53% (from 8.3% to 3.9%) from 1982 to 2015. Wasting prevalence remained stable at low levels (1.8% in 1982 and 1.7% in 2015), whereas overweight increased by 88% (6.5% to 12.2%). Note that the current prevalence of stunting in Pelostas is over twice, and that of underweight over five times higher than the 2.3% that would be expected in a well-nourished population. Details on the interpretation of these prevalence estimates to classify levels of malnutrition in global surveillance are given in Section 13.4.5.

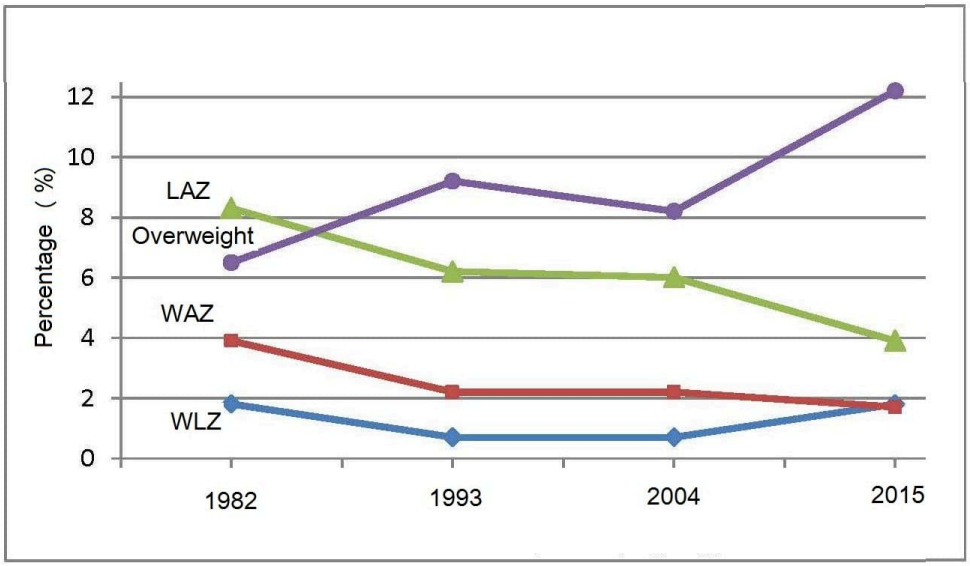


Figure 13.5. Prevalence of malnutrition (length-for-age (LAZ <−2), weight-for-age (WAZ <−2) weight-for-length (WLZ <−2), and overweight at 12mos in the 1982, 1993, 2004, and 2015 birth cohorts in Pelostas, Brazil. Redrawn from Gonçalves et al. (2019).

The WHO indicators stunting, overweight, and wasting are three of the six global nutrition target indicators shown in Box 13.2.

Box 13.2 Global nutrition targets for 2030

- Stunting. Target: 50% reduction in the number of children under 5 who are stunted.
- Anemia. Target: 50% reduction of anemia in women of reproductive age.
- Low birth weight. Target: 30% reduction in low birth-weight.
- Childhood overweight. Target: Reduce and maintain childhood overweight to less than 3%.
- Breastfeeding. Target: Increase the rate of breastfeeding in the first 6 months up to at least 70%.
- Wasting. Target: Reduce and maintain childhood wasting to less than 3%.

From https://www.undp.org/sustainable-development-goals

Indicators for overweight and obesity in both young and older children and adolescents are also based on Z-scores for BMI-for-age of the WHO Child Growth Standard (WHO, 2006) and the WHO growth reference 2007. For children (0–5y) a Z-score for BMI-for-age above +1 is described as being “at risk of overweight”, above +2 as “overweight” and above +3 as “obese”. For children greater than ≥ 5y, de Onis and Lobstein (2010) suggest using the adult classification scheme; children with BMI Z-scores above +1 are described as being “overweight” and above +2 “as obese”.

An important advantage of using Z-scores for population-based applications is that it is valid to calculate the mean and standard deviation for a group of Z-scores. This allows the nutritional status of the entire population to be described. In addition, because the observed Z-scores from a population are often normally distributed, statistical analytical procedures that assume normality such as t-tests and regression methods can be used (Gorstein et al., 1994).

In some circumstances, however, a mean Z-score will not be useful. For example, in some studies of refugees, the death rate among those most severely malnourished can be high and the lower end of the distribution can be truncated: the mean is hardly affected (Yip and Sharp, 1993). In such cases, a comparison of the distribution of indices of the entire population in relation to the reference population is more useful (Figure 13.4). a Z-score of −1.645 corresponds to an observation on the 5th percentile.

Conversely, a value that is at the 85th percentile equates to Z-score of +1.036. The commonly used –3, –2, and –1 Z-scores are the 0.13th, 2.28th, and 15.8th percentiles. An easy way to assess whether the distribution is skewed is to compare the values of the mean and the median. For normal distributions the mean and median will be very similar. As the distribution becomes more skewed, the difference between mean and median increases. Several tests are available for testing normality. Examples include the Kolmogorov-Smirnoff and the Cox tests, of which the latter is generally preferred.

If the distribution of the reference values is normal, percentiles and Z-scores are directly related, as shown in Table 13.2:

Table 13.2. Equivalents of percentile and Z-scores in a normal distribution.

Below mean		Above mean	
Percentile	Z-score	Percentile	Z-score
5.0	–1.645	55.0	0.126
10.0	–1.282	60.0	0.253
15.0	–1.036	65.0	0.385
20.0	–0.842	70.0	0.524
25.0	–0.675	75.0	0.675
30.0	–0.524	80.0	0.842
35.0	–0.385	85.0	1.036
40.0	–0.253	90.0	1.282
45.0	–0.126	95.0	1.645
50.0	–0.000		

The characteristics of the three anthropometric reporting systems — Z-scores, percentiles, and percent-of-median — are compared in Table 13.3. The interpretation of reference limits based on percentiles and Z-scores, is believed to be consistent across all ages and indices. This means that a reference limit of 3rd percentile or –2 Z-scores represents the same degree of malnutrition, irrespective of the anthropometric index used (e.g. weight-for-age or weight-for-height) or the age of the child (Waterlow et al., 1977). Nevertheless, the physiological meaning or consequence of extreme values may differ with age and height.

Table 13.3. The characteristics of three anthropometric data-reporting systems. From: WHO(1995).

Characteristic	Z-score	Percentile	% of median
Adherence to reference distribution	Yes	Yes	No
Linear scale permitting summary statistics	Yes	No	Yes
Uniform criteria across all ages and indices	Yes	Yes	No
Detects changes at extremes of distributions	Yes	No	Yes

Software can be used to compute percentiles and Z-scores for each of the indices. Of the software available, both the manual and the program for WHO AnthroPlus 2009 are in the public domain and may be freely copied, translated, and distributed.

13.1.3 Height-for-age difference

There is some debate about using absolute (centimeters) vs. relative (Z-scores) scales to describe changes in growth in populations of children over time. Leroy and colleagues (2014) argue that although HAZ is useful to assess childen's attained height at a given age, HAZ is inappropriate to evaluate changes in height as children age at both the individual and population level. As shown earlier in the example of the Z-score calculation of an individual, the construction of a Z-score for height-for-age uses as a denominator the Z-score of the reference population. The Z-score is based on cross-sectional data, and reflects the scatter of height values at a specific age. Such Z-scores of the reference population are not constant, however, but actually increase linearly during childhood, thus limiting the usefulness of HAZ to assess changes in height over time.

Instead, Leroy et al. (2014) recommend using height-for-age difference (HAD) to describe and compare height changes as populations of children age. Height-for-age difference is defined as a child's height minus the median height of the growth standard at that age, expressed in centimeters. Leroy and colleagues compared changes in height as populations age based on absolute height-for-age differences (HADs) and HAZs using data from 51 nationwide surveys from low- and middle-income countries (Leroy et al., 2014). A comparison of these changes based on the two methods for children 1–59mo is shown in Figure 13.6.

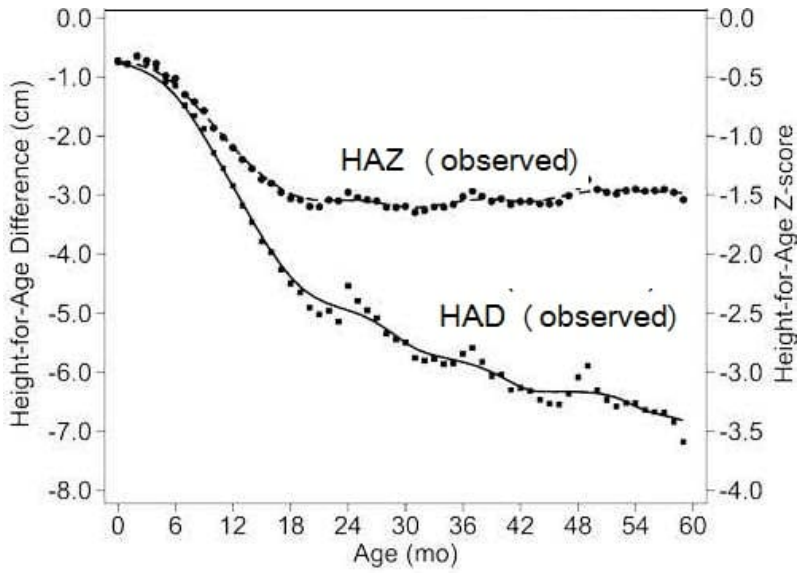


Figure 13.6. Mean HAZ and HAD relative to the WHO standard (1–59mo) by completed month and kernel-weighted local polynomial smoothed values. Data from n = 309,215 children from 51 demo-graphic and health surveys and multiple indicator cluster surveys. HAD, height-for-age difference; HAZ, height-for-age Z-score. Redrawn from Leroy et al.(2014).

In this figure the mean HAZ started below the WHO growth standard at approximately –0.4 Z-scores and fell steadily up to 24mo, after which it stabilized and increased only slightly.

Based on the mean HAD curve, the children started with an average height deficit of 0.8cm, with the most pronounced faltering (i.e., steepest slope) evident between 6 and 18mo, as shown for the HAZ curve. Nevertheless, in contrast with the trend in the HAZ curve, the deficits in linear growth depicted by the HAD curves continued to increase after 18mo, albeit at a lower rate than before 18mo. Indeed, the slopes of the HAD curve provide no indication that the process of growth faltering leveled off even at 5y. The bumps noted in the curves just after 24, 36, and 48mo are reportedly due to the tendency to report age in completed years rather than exact months (Victora et al., 2010). Such growth deficits were also evident in five regions (E. Europe and Central Asia, North Africa and Middle-East, Latin America and Carribean, Africa South of the Sahara, South Asia), although the magnitude of the deficits varied between regions. See Leroy et al. (2014) for more details.

Globally, 70% of the absolute deficit accumulated in height (HAD) at 60mo could be attributed to faltering during the first "1000" days (conception to 24mos), but 30% was due to continued increases in deficit from age 2–5y. However, such continued increases were masked by the use of HAZ with the age-related changes in Z-scores. These findings emphasize the importance of using the mean HAD to provide an accurate assessment of population-level catch-up in linear growth at all ages. For more discussion see Leroy et al. (2014; 2015) and Lundeen et al. (2014).

13.1.4 Percent-of-median

Growth indices can be expressed as percent-of-median value of the reference data when the distribution around the median value is unknown. Percent-of-median is the ratio of a measured anthropometric value (e.g. weight) in the individual, to the median value of the reference data for the same age or height, expressed as a percentage. It is especially useful when the distribution of the reference data has not been normalized, such as the earlier Harvard reference data, and explains why so many of the first classification schemes based on the Harvard reference data (e.g., Gomez and Wellcome classifications) used percent-of-median (Gorstein et al., 1994).

Use of percent-of-median does not provide the same information as the Z-score, and the relationship between these two criteria differs with age and height. Such discrepancies may be especially important if intervention activities are prioritized across populations according to the prevalence of low anthropometric indices. A comparison of selected Z-scores and percent-of-median curves for weight-for-height, height-for-age, and weight-for-age is given in Gorstein et al. (1994).

A further limitation of the percent-of-median is that, unlike percentiles or Z-scores, the interpretation of specific percent-of-median varies across age groups and growth indices Table 13.3. This arises because the calculation of the percent-of-median does not take into account the distribution of the data within the reference set, and particularly the differing widths of the distributions of the weight-for-age, weight-for-height, and height-for-age indices (Waterlow et al., 1977). Indeed, it is this variability that inhibits the universal use of a constant percentage of the reference median (e.g. 70%) across all ages and for all growth indices. For example, 60% of median weight-for-age represents a more severe state of malnutrition for younger than for older children. Moreover, when the index weight-for-height is used, 60% of the median is inappropriate; such a deficit at any age is incompatible with life (Dibley et al., 1987).

This page titled 11.2: Anthropometric indices and their modes of expression (13.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

11.3: Use of anthropometric indices in clinical settings (13.2)

Anthropometry in clinical settings has the following applications for individuals: screening for early detection of abnormal changes in growth, and assessing the response of an individual to therapy (WHO, 1995). Increasingly, the three conventional growth indices (weight-for-age, length/height-for-age, and weight-for-length/height) are used together with BMI for children age 0–19y, all expressed as either percentiles or Z-scores. Recumbent length is the recommended measurement for children younger than 24mo, whereas in those aged ≥ 24mo, standing height should be used.

Health care providers who measure and assess the growth of children or supervise these activities can access the WHO Training Course on Child Growth Assessment which provides instructions on both measuring a child's growth and interpreting growth indicators manually using the WHO international growth reference charts. The course is available from (WHO).

Alernatively, software can be used to compute percentiles or Z-scores for the three growth indices (weight-for-age; length/height-for-age; and weight-for-length/height) and BMI-for age for a child aged from 0–19y using two modules — Anthropometric calculator and Individual assessment — both available in WHO AnthroPlus 2009. Calculations are based on the WHO Child Growth Standard for children 0–5y (WHO, 2006) and the WHO Growth Reference for children aged 5–19y (de Onis et al., 2007).

13.2.1 Screening to identify abnormal growth

Accurate measurements of weight and length or height on each child are needed. WHO has developed an Anthropometry Training Video which can be downloaded from WHO. Box 13.3 summarizes some of the points emphasized by WHO to ensure the collection of accurate growth data.

Box 13.3 Points to consider to ensure accurate growth measurements

- Make sure all equipment is correctly calibrated on a regular basis.
- Conduct training based on recommended measurement protocols as well as standardization sessions for those who collect the data.
- Take the child's date of birth from a written record if available. Otherwise ask for both the child's date of birth and age on the day measured, since the year of birth is frequently reported incorrectly. If birth dates are not recorded or known with certainty, probe the caregiver for the approximate date of birth based on a local events calendar.
- If WHO AnthroPlus is used, the software calculates the precise age, depending on the age of the child in days (for those aged 0–60mos) or months (for those aged 5–19y). Alternatively, secure the WHO child age calculator consisting of a rotating disk mounted on a calendar in polyvinyl chloride film, if attending a training workshop in a regional WHO Regional Office.
- Measure and record recumbent length in children younger than 24mo or standing height from 24mo onwards, where possible. There is a 0.7cm difference between length and height. If a child older than 2y is measured lying down instead of standing, then 0.7cm is subtracted by the software to derive an estimated height. Alternatively, if a child younger than 24mo is measured standing, then 0.7cm is added by the software to derive an estimated length.
- Record whether the child has edema or not before measuring the child's weight: edema is a clinical sign of severe undernutrition that can mask a very low weight. To check for the presence of edema, grasp the child's foot so that it rests in your hands with your thumb on top of the foot. Press the thumb down gently for a few seconds. If a pit (dent) remains in the foot after lifting your thumb, then the child has edema. No weight measurement should be taken if edema is present.
- Once the data on age, sex, weight, and length/height have been recorded, check the results and if the child appears to have extreme Z-score values, the child should be re-measured immediately. Note: calculating the Z-score values before proceeding with the measurement of the next subject is good practice.

From: Child Growth Training Manual. (WHO).

Table 13.4. summarizes examples of growth indicators based on Z-scores relative to the WHO Growth References which have been shown to reflect growth problems. More insight on the nature of the growth problems of a child can be gained by plotting

and examining all the growth measurements together on the WHO growth charts.

Table 13.4. Interpretation of growth indicators based on Z-scores indicative of growth problems. From: (WHO Training Course on Chi

Indicators. Notes:
1. A child in this range is very tall. Refer a child in this range for further assessment if you suspect an endocrine disorder (e.g., if pare
2. A child whose weight-for-age falls in this range may have a growth problem, but this is better assessed from weight-for-length/he

3. A child above 1 shows possible risk. A trend towards a Z-score of + 2 shows definite risk.

4. It is possible for a stunted or severely stunted child to become overweight.

Z-score	Length/height for age	Weight for age	Weight for length/height
Above 3	See note 1	See note 2	Obese
Above 2			Overweight
Above 1			Possible risk of overweight (See note 3)
0 median			
Below –1			
Below –2	Stunted (See note 4)	Underweight	Wasted
Below –3	Severely stunted (See note 4)	Severely underweight	Severely wasted

13.2.2 Screening to identify abnormal changes in growth

In general, to identify abnormal changes in growth, measurements in a child should be repeated at monthly intervals during the first 6mo, then every 2–3mo during later infancy and early childhood, and then every 6–12mo(Guo et al., 1991). Special attention must be given when measuring length in children below 2 years of age; guidelines are available in FANTA Anthropometric Guide (2018).

Such serial measurements provide information on the pattern of growth over time. This enables a child with failure to thrive to be identified. Abnormal growth is apparent from the direction of the growth curve: an abnormal height-for-age growth curve is horizontal or, in the case of weight-for-height, moves downwards. The former indicates that the child is not growing, and the latter shows that the child is losing weight. Such a trend may indicate that the child has been ill. Alternatively, the trend may be desirable if the child is overweight. Consult the WHO Training Course on Child Growth Assessment for more examples.

As an example, Figure 13.7 depicts diagrammatically the growth history of a 7y girl who was following a macrobiotic diet, and whose growth, based on height-for-age, was measured and plotted at least annually on a distance growth chart. Curves for –2 Z-score, the median, and +2 Z-score for the growth reference for height-for-age are shown. Growth indices that fall below –2 Z-score or above +2 Z-score are defined as abnormal. In Figure 13.7 the child's height-for-age growth curve fell below the –2 Z-score reference at 2y, at which point the curve became nearly horizontal, indicating that the child was growing very, very slowly.

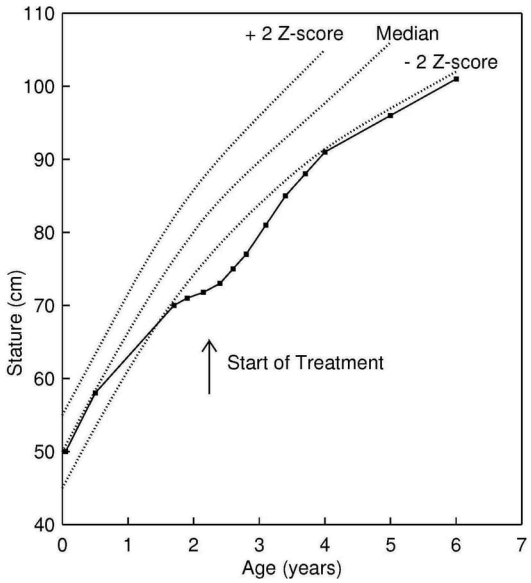


Figure 13.7. Diagrammatic growth history of a 7y girl following a macrobiotic diet.

When a child exhibits deceleration or slow growth of length or height, as shown in Figure 13.7, then the child must be evaluated to distinguish between a shifting of linear growth, sometimes termed “rechanneling”, constitutional growth delay, and underlying pathology. Any shift in linear growth is determined by genetic influences, and tends to occur within the first 18mos of life, when the growth of an apparently healthy child may cross one or more Z-score or percentile lines. In contrast, the exact cause of constitutional growth delay is unknown. It occurs later than rechanneling, and is characterized by a temporary delay in skeletal growth and height of a child, and is not accompanied by any other physical abnormalities causing the delay. Alternatively, slow growth can arise from many pathological conditions, such as malnutrition (Figure 13.7), chronic disease (e.g., tuberculosis), non-organic failure to thrive (caused by a severe psychosocial disturbance in the family), or organic disorders. The latter may encompass congenital and acquired disorders, or abnormal function of the endocrine system or metabolism, and gastrointestinal dysfunction (e.g., malabsorption); details are given in WHO (1995). Guidelines on how to investigate causes of undernutrition in a child are given in the WHO Child Growth Training Course described earlier.

Although changes in a child's growth are apparent on distance growth charts, abnormal changes in the rate of growth of a child can be detected much earlier when growth velocity charts, rather than distance growth charts, are used: growth velocity measurements are much more sensitive to recent changes in growth but require careful measurement procedures to be of value. WHO has developed a set of growth velocity charts which are recommended for international use (de Onis et al., 2011).

The prevalence of overweight and obesity is increasing globally in children younger than 5y, and is an important contributor to diabetes and other noncommunical diseases in later life. The WHO Child Growth Training Course provides guidelines on how to investigate causes of overweight and obesity in infants and children 0–6mo; 6mo–2y; and 2–5y.

Trends in pediatric overweight and obesity in the United States are an ongoing serious health concern as about 17% of U.S. children and adolescents are obese. In the U.S., the use of BMI in association with the CDC (2000) BMI percentiles are recommended by the U.S. Endocrine Society to diagnose overweight or obesity in children and adolescents > 2y. Overweight is diagnosed in the U.S. if a child or adolescent > 2y has a BMI > 85^{th}percentile but < 95th percentile for age, as obese if the BMI is > 95th percentile, and as extremely obese if the BMI is > 120% of the 95th percentile or > 35kg/m^2 (Styne et al., 2017). They recommend calculating, plotting, and reviewing a child's or adolescent's BMI percentile at least annually during well-child and/or sick visits, and if a child or adolescent has a BMI > 85th percentile, then they should be evaluated for potential comorbidities. The Endocrine Society also provides advice to prevent obesity; for more details see Styne et al. (2017).

13.2.3 Assessing response to therapy

Serial measurements are also useful when assessing the response to therapy, as shown in Figure 13.7; again at least two anthro-

pometric measurements are required. Selection of the indicator of response to the therapy must be made carefully. It must take into account the possible time lag between the start of the intervention and the time when a response is apparent.

In the example shown in Figure 13.7, the child was enrolled in a 2y intervention program because at age 2y the child's length-for-age growth curve fell below the –2 Z-score reference, at which point the curve tended to flatten out, indicating that the child was growing slowly. After 2 years of treatment, catch-up growth in response to the therapy is clear, although the child did not approach the –2 Z-score reference until she was 4y.

An infant with severe acute malnutrition (SAM) is known to respond to a nutrition intervention by first putting on weight, only later "catching" up in linear growth. Consequently, the recommended discharge criteria from a therapeutic feeding program for a child with SAM defined by a weight-for-height < –3 Z-score (based on the WHO Child Growth Standard) is when the child has reached a weight-for-height above –2 Z-score and no edema (WHO/UNICEF, 2009). These discharge criteria are based on the lower risks of mortality compared to those children below –3 Z-score, as shown in Figure 13.8. If individuals do not respond to the therapy, then some other medical treatment may be necessary to improve the poor growth, and the earlier therapy should be discontinued.

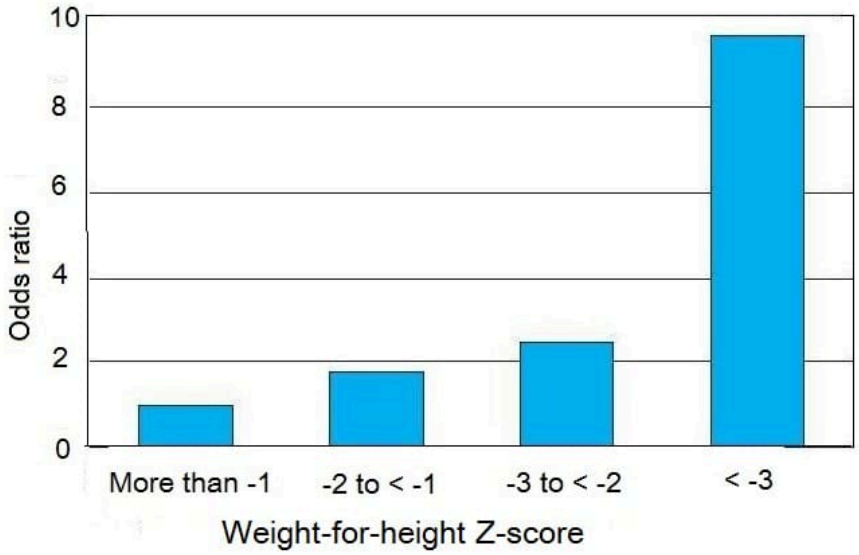


Figure 13.8. Odds ratio for mortality by weight-for-height. Note: reference category: children with a weight-for-height > –1 SD. From WHO Identifying Severe Acute Malnutrition. (WHO/UNICEF, 2009).

This page titled 11.3: Use of anthropometric indices in clinical settings (13.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

11.4: Screening systems to identify individuals at risk in a population (13.3)

Anthropometry has a range of applications that can be used at the population level. Increasingly, screening systems are being used to map countries according to levels of severity of malnutrition(UNICEF/WHO/World Bank, 2021), and to identify priority countries for action. Screening systems are also being used to monitor progress across countries in an effort to mobilize countries to achieve the Sustainable Development Goals.

In public health, it is often not possible to obtain serial measurements on the same individuals to identify abnormal changes in growth, as described in the clinical setting in Section 13.2.2. Instead, screening systems are used to identify individuals in a population "at risk" to malnutrition and who may require intervention. Some screening systems are used only in emergency settings, when low-cost portable equipment and the use of trained unskilled personnel are primary considerations. In such circumstances, only one anthropmetric measurement (e.g., MUAC) may be taken. All the systems use at least one anthropometric meaurement and one or more reference limits drawn from appropriate reference data. In some cases, cutoff points based on functional impairment or clinical signs of deficiency or mortality risk may be used.

An optimal screening system is one in which no misclassification occurs; all the individuals designated as "at risk" are actually malnourished (i.e. there are no false positives; sensitivity is 100%). Similarly, those individuals classified as "not at risk" are truly unaffected (i.e. there are no false negatives; specificity is 100%).

In practice, classification schemes are not perfect; some misclassification will always occur so that some individuals identified as "at risk" to malnutrition will not be truly malnourished (false positives), and others classified as "not at risk" to malnutrition will in fact be malnourished (false negatives). Unfortunately, specificity and sensitivity data for the anthropometric indices selected are usually not known for the target population. Instead, when selecting a screening system, values for sensitivity and specificity obtained elsewhere are often used and assumed to be appropriate for the target population (Habicht et al., 1979; 1982).

Unfortunately, literature values for sensitivity and specificity are unlikely to apply to the target population when the indices assess an outcome (e.g. low birth weight) which is influenced by several biological factors. For example, low birth weight can be due to both prematurity and intrauterine growth retardation, and its sensitivity as an indicator of neonatal mortality is greater for the former than the latter (WHO, 1995). Factors such as the severity of the deficiency, age, and sex may also affect the magnitude of the expected response of the chosen indicator. A detailed discussion of the selection criteria for indicators can be found in Habicht et al. (1982) and Habicht and Pelletier (1990).

A more general way of selecting an optimal screening system is shown in Figure 13.9, that depicts the relationship between the sensitivity and specificity for both height-for-age and weight-for-height in predicting 2y survival. The position of the Receiver-Operating Characteristic (ROC) curves indicates that, in this case, height-for-age is the better index. For more details on the use of ROC curves, see Chapter 15. Brownie et al.(1986) also provide details of statistical methods that can be used to select the best indicator.

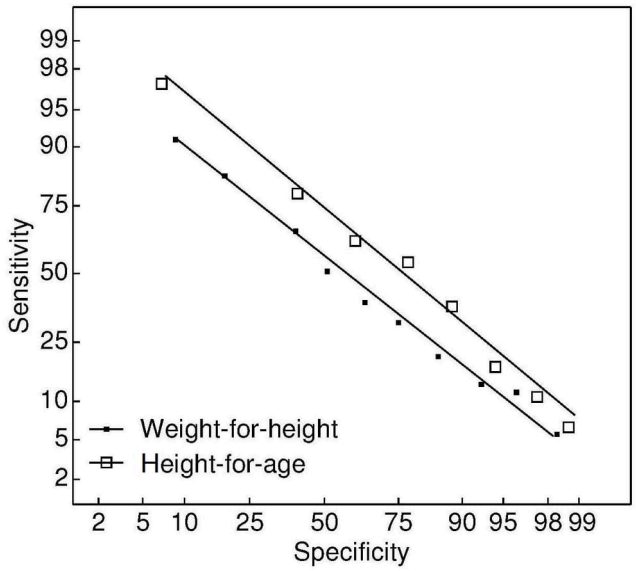


Figure 13.9. Relationship between sensitivity (%) and specificity (%) for height-for-age and weight-for-height in predicting 2y survival. The Receiver-Operating Characteristic (ROC) plots have been linearized by the Z-transformation. The linearized height-for-age ROC plot lies above the weight-for-height ROC plot, indicating that the former index generates fewer errors of classification at every level of sensitivity and specificity. Redrawn from Habicht et al. (1982).

In many of the early classification systems, cutoff points were defined as a designated percentage of the median of the reference population, for the reason outlined earlier. Now the approach preferred by WHO is to use reference limits based on Z-scores (WHO, 1995). Examples of screening systems that use this approach and are based on a single anthropometric index or a combination, and that can be used to classify individuals in a population "at risk" to malnutrition are outlined below.

13.3.1 Use of weight-for-length/height for young children

As noted earlier, low weight-for-length/height is described as "thinness" and reflects a pathological process referred to as "wasting". It arises from a failure to gain sufficient weight relative to length or height, or from losing weight. The prevalence of wasting in a population is defined as the proportion of children with a Z-score for weight-for-length or height less than –2 (i.e., < 2SDs below the age- and sex-specific WHO reference mean). Wasting is said to represent acute malnutrition which should be treated with an acute intervention to reverse the wasting and prevent death. The onset of wasting is often highly seasonal with an onset that is fairly acute and often associated with both changes in the food supply and the prevalence of infectious diseases. The highest prevalence of low weight-for-length often occurs during the postweaning period(12–23mos) (WHO, 1986).

WHO further classify low weight-for-length/height as:"moderate acute malnutrition" when the weight-for-length/height is between –2 and –3 Z-score, and "severe acute malnutrition" (SAM) if the Z-score is below –3 of the WHO Child Growth Standards (WHO/UNICEF, 2009). Treatment of children with moderate acute malnutrition involves improving their existing diets by nutritional counseling to ensure consumption of nutrient-dense foods that meet their additional needs for weight and length gain and functional recovery. Only in settings where there is a high prevalence of wasting or food insecurity at the community or household level should food supplements, usually fortified blended flours, be provided (WHO/UNICEF/WFP/UNHCR, 2009). For more details on the nutrient needs of children with moderate acute malnutrition, see Golden (1990).

Currently, children with SAM (weight-for-length/height Z-score below –3) are treated most frequently with Ready-to-Use Therapeutic Foods or F75 and F100 milk-based diets; for more treatment details see WHO (2013). WHO recommends that all children with SAM who also have medical complications, severe edema, poor appetite, or present with one or more Integrated Management of Childhood Illness danger signs, should be admitted to hospital for inpatient care. Transfer of children with SAM from inpatient to outpatient care should be based on the clinical condition of the children and not on specific anthropometric

outcomes. Justification for the use of weight-for-length/height Z-score below –3 for identifying SAM is shown in Box 13.4.

Box 13.4. Justification for use of weight-for-height Z-score below - 3, for identifying SAM

- Risk of death for children with a weight-for-height below –3 Z-score is much higher than for children with weight-for-height Z-score above –3 Z-score (see Figure 13.8);
- Children with weight-for-height Z-score below –3 Z-score have a higher weight gain when receiving a therapeutic diet compared to other diets, resulting in faster recovery;
- There are very few children below –3 Z-score (i.e., < 1%) in a well-nourished population;
- Use of the recommended protocols and appropriate therapeutic foods is not associated with any known risk or negative effects on the children.

Modified from WHO/UNICEF (2009).

WHO provide weight-for-length reference cards for children less than 87cm, and 87cm and above, by sex, that can be used to assess the Z-score of a child in the field once their weight (kg) and height (cm) have been measured. These cards are available from WHO.

As noted earlier, the WHO recommended criterion used for discharge when weight-for-height has been used as an admission criterion is when weight-for-height reaches > –2 Z-score with no pitting edema (WHO, 2013). This discharge criterion is based on the lower risks of mortality compared to those children below –3 Z-score as shown in Figure 13.8. Percentage weight gain should no longer be used as a discharge criterion (WHO, 2013).

Note that use of the new WHO Child Growth Standard in low income countries has resulted in a 2–4 times increase in the number of infants and children falling below –3 Z-score compared to using the former NCHS / WHO reference population published by WHO in 1983.

Overweight in young children is becoming increasingly common, as noted earlier. Although WHO defines any child as overweight with a weight-for-length/height Z-score above +2, and as obese with a weight-for-length/height Z-score above +3, increasingly BMI-for-age Z-scores are being used to define risk of overweight, overweight, and obesity in children.

13.3.2 Use of MUAC for children

Mid-upper-arm circumference (MUAC) is used for screening individual children for targeting interventions in the community when weight and stature measurements are impossible and the precise age of the child is unknown. In such cases, a fixed cutoff point for MUAC is often used. For example, to diagnose children with SAM aged 6–60mo, WHO (2009) recommends a fixed MUAC cut-off point of 115mm as one of the independent criteria that can be used by trained community health workers and community members. This MUAC cutoff was chosen because children with a MUAC less than 115mm were observed to have a highly elevated risk of death compared to those with a MUAC above 115mm (Myatt et al., 2006). In addition, the prevalence of SAM based on a MUAC cutoff of 115mm and that defined by a weight-for-height below –3 Z-score based on the new WHO Child Growth Standard is believed to be similar, based on an analysis of measurements on more than 450,000 children aged 6–60mo in 31 countries.

Table 13.5. summarizes the WHO diagnostic criteria for SAM in children aged 6–60mos. To identify children in the community with SAM for treatment, WHO recommends the measurement of MUAC together, where possible, with examination for bilateral pitting edema as the preferred dignostic criteria. Provision to mothers of MUAC tapes with a cut-off value of 115mm have been used in an effort to detect malnutrition early before the onset of complications, and thus reduce the need for inpatient treatment (Blackwell et al., 2015).

Table 13.5. Diagnostic criteria for severe acute malnutrition in children aged 6–60mos. From: WHO/UNICEF, 2009.

[1] Based on Child Growth Standards (WHO).

[2,3] Independent indicators of Severe Acute Malnutrition that require urgent action. From:WHO/UNICEF, 2009.

Indicator	Measure	Cut-off
Severe wasting[2]	Weight-for-height[1]	< –3 Z-score
Severe wasting[2]	MUAC	< 115mm
Bilateral edema[3]	Clinical sign	

The WHO recommended discharge criteria from treatment for children diagnosed with SAM using the MUAC cutoff is based on a MUAC > 125mm and no edema for at least 2wk. If children have been admitted based only on bilateral pitting edema, then the criteria for discharge should be the one used in programs. To avoid any relapse, children with SAM who have been discharged from treatment should be monitored periodically (WHO, 2013).

Concerns have been raised over the use of the MUAC cutoff alone as a prognostic indicator of mortality. As an example, Grellety and Golden (2018) claim that up to 45% of SAM children at high risk of death will not be identified using this approach. This finding was based on an examination of the relative mortality rates of children 6–60mos who had SAM (n=76,887) classified by the three diagnostic criteria depicted in Table 13.5. Instead, these investigators urge that the use of weight-for-height Z-scores should be retained as an independent criterion for diagnosis of SAM, and innovative methods developed to identify those children with a low weight-for-height Z-score but not a low MUAC in community screening programs. In the interim, they suggest that use of both a weight-for-height < –3 Z-score and MUAC < 115mm should be retained and used routinely as diagnostic criteria for SAM in an effort to identify those children who should receive treatment.

Use of a fixed cutoff assumes that MUAC is relatively independent of age for children 1–5y (Burgess and Burgess, 1969). This assumption has been questionned by several investigators. Data from both affluent and nonaffluent populations show that MUAC is age dependent, resulting in over-diagnosis of wasting among younger children and under-diagnosis among older ones, when a fixed cutoff point is used (de Onis et al., 1997). In view of these concerns, WHO has developed MUAC-for-age reference data for children 6–60mo (Figure 13.10). The curves show important age-specific differences, as well as significant sex-specific differences for boys and girls < 24mo. In the field, the use of MUAC-for-age reference data is no more difficult than using weight-for-height, the conventional index of wasting.

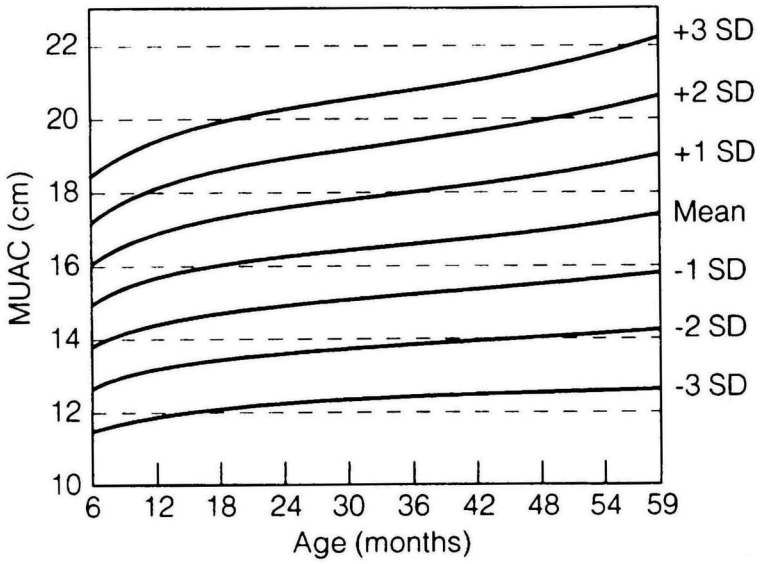


Figure 13.10. Mid-upper-arm circumference (MUAC) for age growth reference curves for boys 6–59mos. Redawn from: de Onis et al. (1997).

13.3.3 Use of MUAC and Calf Circumference for adults

The importance of adult undernutrition is becoming increasingly recognized. Both the Food and Agricultural Organization (FAO) and WHO have adopted BMI to assess adult undernutrition and have defined cutoffs grading the severity of low BMI in adults. However, in emergency situations, measurements of weight and height, required to calculate BMI, are difficult to obtain from severely emaciated individuals who often cannot stand unaided. Therefore, MUAC and calf circumference, measurements that only require a tape measure, have been investigated as a substitute for BMI.

Significant correlations between measurements of MUAC and BMI have been demonstrated in women of reproductive age (Khadivadeh, 2002),and severely under-nourished adult populations from low-income countries (Collins, 1996). Reports of positive associations between calf circumference and BMI have been primarily reported in clinical settings and mostly in the elderly (Aparecida et al., 2012). However, there is little known about the relationships between both MUAC and calf circumference (CC) and health outcomes in adults.

WHO has proposed universal cutoff values for MUAC of 22cm for females and 23cm for males, and below 31cm for calf circumference (for both sexes) to screen for undernutrition (WHO, 1995). However, in view of the reported regional and ethnic variations in body composition, these cutoff values may not be appropriate for Asian populations who have a higher proportion of body fat than other races (e.g., Caucasians and Africans) (Deurenberg et al., 1998) (see Chapter 11).

As an example, in a study in northern Vietnam of women of reproductive age (n=4,981), a higher MUAC cutoff (23.5cm) was used to screen for underweight in comparison to the BMI. This 23.5cm MUAC cutoff had a sensitivity of 89.1%, a specificity of 71%, and a positive predictive value of 84% when analyzed using ROC curves. If the lower WHO cutoff (22cm) had been used, many women at risk for undernutrition would have been misclassified as healthy. By contrast, the calf circumference cutoff proposed by WHO (31cm) was found to be appropriate, although slightly less accurate than the MUAC cutoff for predicting underweight (Nguyen et al., 2014). More studies are needed to examine the utility of MUAC and calf circumference in predicting functional outcomes.

MUAC during pregnancy has also been investigated as a screening tool for risk of low birth weight and late fetal and infant mortality. Maternal arm circumference is relatively stable during pregnancy, so its measurement is independent of gestational age (Mohanty et al., 2006). Several investigators have reported that MUAC, when used as a proxy of nutritional status of women during pregnancy, is a strong predictor of maternal mortality (Christian et al., 2008) and fetal outcomes such as low birth weight (Mohanty et al., 2006; Assefa et al., 2012). In contrast, calf circumference is not recommended as a proxy measurement of status during pregnancy. Edema is increasingly common as pregnancy advances, so calf circumference may be increased particularly in late pregnancy (WHO, 1995).

13.3.4 MUAC-for-height: QUAC stick

The Quaker arm circumference measuring stick (QUAC stick) was developed as a rapid, cheap, and simple screening tool for the nutritional assessment of children. It avoids the use of cumbersome scales and minimizes errors in measuring height (Arnhold, 1969). The measuring stick is used to compare a child's MUAC with the two reference limits of the MUAC reference data corresponding to the child's height.

The modified QUAC stick developed by WHO uses MUAC-for-height data from the MUAC-for-age reference for children < 5y (de Onis et al., 1997; Mei et al., 1997). Three sets of MUAC-for-height curves were developed: one for boys, one for girls, and one for both boys and girls combined. The curves are for children of height 65–145cm. Recumbent length measurements were adjusted to standing height when the MUAC-for-height reference data were constructed to eliminate any discrepancy caused by differences in the two stature measurements.

Mei et al. (1997) also provide instructions on the construction and use of the modified QUAC stick shown in Figure 13.11 that uses the WHO MUAC-for-height reference data. Construction of two sticks, one for the boys and one for girls, is recommended.

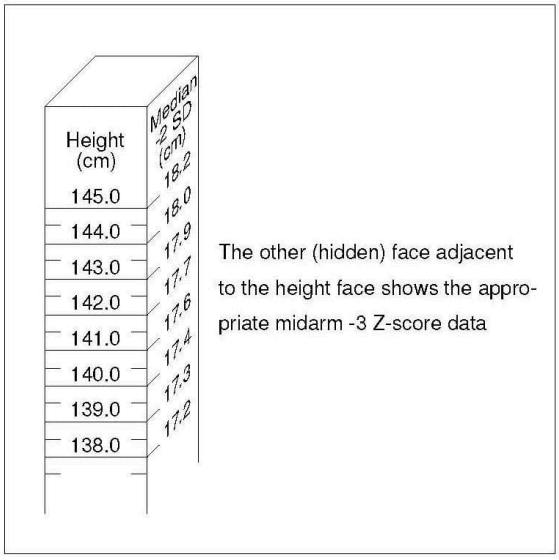


Figure 13.11. The WHO-modified QUAC stick. Redrawn from: Mei et al. (1997).

The WHO-modified QUAC stick consists of a vertical stick, about 150cm long and 3 × 3cm in cross-section. One face is marked from 0 at the bottom up to 145cm near the top in 0.5cm increments. On one adjacent face, the corresponding reference values for the median –2 Z-score of MUAC-for-height are marked; the other face, adjacent to the height measurements, shows the values for median –3 Z-score of the MUAC-for-height.

To use the QUAC stick, it is first placed firmly upright on a platform against a vertical wall. The child is then asked to stand straight with his or her back against the height measure. Next, the MUAC of the child is measured using the technique described in Chapter 11. Then a note is made about whether the value is below that given on the “median –3 Z-score” face at the point on the stick marked by the child's height. If the measurements is below this value, the child is noted as having a severe nutritional deficit; if above and yet still below the appropriate value on the “median –2 Z-score” face, the deficit is recorded as moderate. In this way, all children are identified as having severe, moderate, or no deficit. Note that the method eliminates the necessity of recording both the height and MUAC of the child for subsequent comparison with reference data.

The performance of MUAC-for-height for screening was evaluated using weight-for-age < –2 Z-score as the “gold standard” for identifying malnourished children (Mei et al., 1997). Results indicated that the sensitivity and specificity for MUAC-for-height was better than MUAC based on a fixed cutoff for screening malnourished children 6–59mo from Sri Lanka, Nepal, Togo, and Malawi.

13.3.5 Weight-for-height wall chart

A wall chart developed by Nabarro and McNab (1980) is based on weight-for-height and three percentage ranges of the corresponding median weight-for-height of the earlier NCHS reference data. These are 70%–80%, 80%–90%, and 90%–110%. Wasted children are identified as those with a percentage weight-for-height less than 80% of the reference median. The chart has the advantage of visually identifying the extent of wasting in a very simple manner. In addition, during nutritional rehabilitation, a target weight can be readily identified.

The chart is fixed to a smooth wall, and the child is positioned on a level floor against the chart at a point corresponding to the child's weight (Figure 13.12).

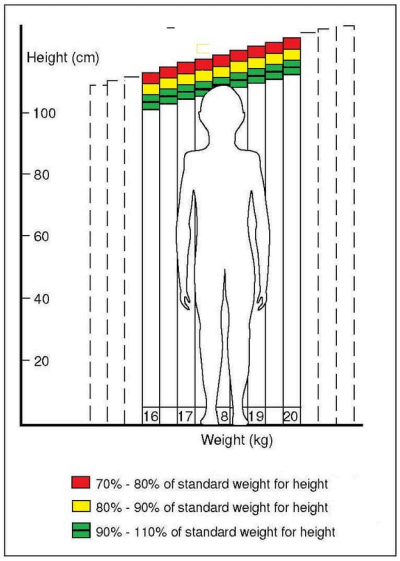


Figure 13.12. Weight-for-height wall chart. Redrawn from: Nabarro and McNab (1980).

The top of the child's head will fall into one of three color-coded categories on the chart, equivalent to 70%–80% (red), 80%–90%(yellow), and 90%–110% (green) of the expected weight-for-height and representing various degrees of malnutrition. The color should be read off the chart at the point at which the examiner's fingers touch the column marked with the child's weight.

For small children, two persons may be required to position the child correctly for the measurement. To avoid misclassifying children with edema or ascites (fluid in the abdominal cavity), a brief clinical examination should also be undertaken. The charts can be modified to include different colors, numbers of colors, and alternative growth reference data and reference limits, depending on the resources and needs of the population.

13.3.6 Use of Body Mass Index for children

In view of the increasing worldwide prevalence in under-five children of overweight and obesity, and their subsequent role in diabetes and other chronic diseases in adulthood, WHO has endorsed “reduce and maintain childhood overweight to less than 3%” as one of the six global nutrition targets (Box 13.1; WHO, 2014). The worldwide prevalence of overweight and obesity among preschool children has increased from 4.2% (95% CI:3.2%, 5.2%) in 1990, to 6.7% (95% CI: 5.6%. 7.7%) in 2010, and is expected to have reached 9.1% (95% CI:7.3%, 10.9%), or about 60 million in 2020 (de Onis & Lobstein, 2010).

Body mass index (BMI) is considered the most practical, universally applicable, and non-invasive index of overweight and obesity. An increased BMI-for-age in childhood and adolescence is associated with higher percentages of body fat and known risk factors for cardiovascular disease.

The WHO defines a child 0–5y as “at risk of overweight” if BMI Z-score is > +1 and < +2; “overweight” if BMI Z-score > +2 and “obese” if BMI Z-score > +3, based on the WHO Child Growth Standard (WHO, 2006; de Onis & Lobstein, 2010). For children aged 5–19y, BMI-for-age Z scores above +1 and above +2 based on the WHO 2007 growth reference data (de Onis et al., 2007) are applied to define overweight and obesity, respectively.

In the United States, use of BMI in association with the CDC (2000) BMI percentiles are recommended by the U.S. Endocrine Society to diagnose overweight or obesity in children and adolescents > 2y (Kuczmarski et al., 2000), as noted earlier (Section 13.2.2); (Figure 13.13). The BMI-for-age percentiles (5th, 10th, 25th, 50th, 75th, 90th, and 95th) are available for children aged 2–20y. The weight data for children > 6y who participated in the U.S. NHANES III survey were excluded from these BMI growth charts because the inclusion of these data shifted the upper percentile curves. Consequently, the CDC (2000) growth reference is not a purely descriptive growth reference because it does not contain representative data for the BMI variables.

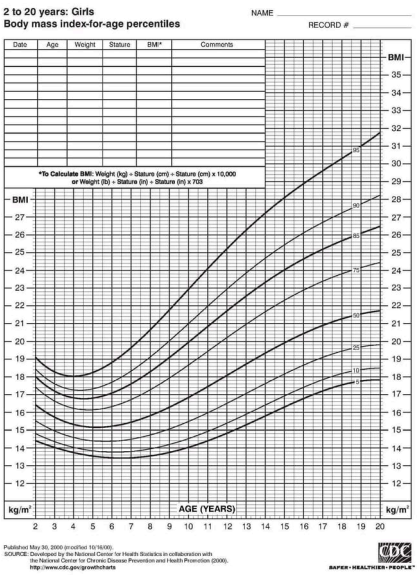


Figure 13.13. CDC 2000 growth chart. BMI-for-age percentiles: girls 2–20y. From Kuczmarski et al. (2000).

13.3.7 Use of Body Mass Index for adults

Over the past few decades both low- and middle-income countries have faced a double burden of malnutrition as a result of the rapid increase in the prevalence of overweight together with a persistent prevalence of underweight. For example, in 2016, the global estimate among adults for overweight or obesity was more than 33% while underweight was 10%(NCD Risk Factor Collaboration, 2017). Both conditions are associated with increased morbidity and mortality and increase the risk of developing non-communicable diseases (Zhang et al., 2021; NCD Risk Factor Collaboration, 2017). This increasing trend for a double burden of malnutrition has been attributed to unequal changes in food production, dietary patterns, and physical activity as communities have encountered the nutrition transition (Hoque et al., 2015).

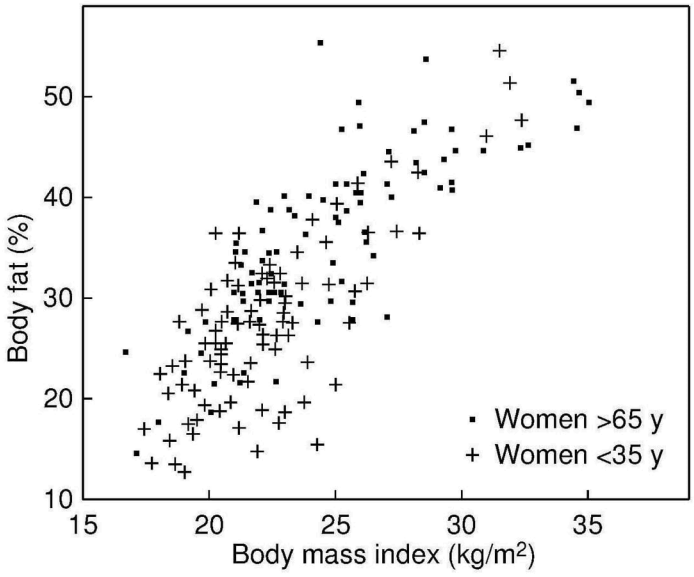


Figure 13.14. The relationship between body mass index and percentage of body fat in both younger and older women from New York City. Body fat determinations based on a four-compartment body composition model. Redrawn using data from Gallagher et al.(1996).

Body mass index (BMI) is the most widely used index of overweight and obesity in adults as well as children. Body mass index is relatively unbiased by height and generally correlates with laboratory-based measures of adiposity in most younger and older adults, as shown in (Figure 13.14).

WHO has recommended the use of a graded classification of overweight and obesity to:

- Compare weight status within and between populations
- Identify individuals and groups who are at increased risk of morbidity and mortality
- Identify priorities for intervention at both the individual and community levels
- Allow interventions to be evaluated appropriately

Table 13.6 shows the classification for underweight as well as the graded classification of overweight and obesity in adults based on BMI and recommended for international use by WHO (2000).

WHO (2000) also recommend that the optimum population mean BMI should be between 21.0 and 23.0kg/m^2. The classfication in Table 13.6 is based primarily on the association between BMI and mortality, after taking into account the confounding effects of cigarette smoking and co-existing chronic disease.

The general term “overweight” is used by WHO (2000) to describe all individuals with BMI ≥ 25kg/m^2. Individuals with BMI from 25.0 to 29.9 are further termed “pre-obese”. Three classes of obesity are defined based on BMI cutoffs of 30, 35, and 40kg/m^2. This classification system is not intended for use with pregnant and lactating women, or persons < 18y.

Although WHO recommends the universal use of these cutoffs internationally (WHO Expert Consultation, 2004), nevertheless WHO recognized that a substantial proportion of Asians were at high risk of type 2 diabetes and cardiovascular disease with BMIs lower than the existing WHO cutoff for overweight (i.e., ≥ 25kg/m^2). (Figure 13.15).

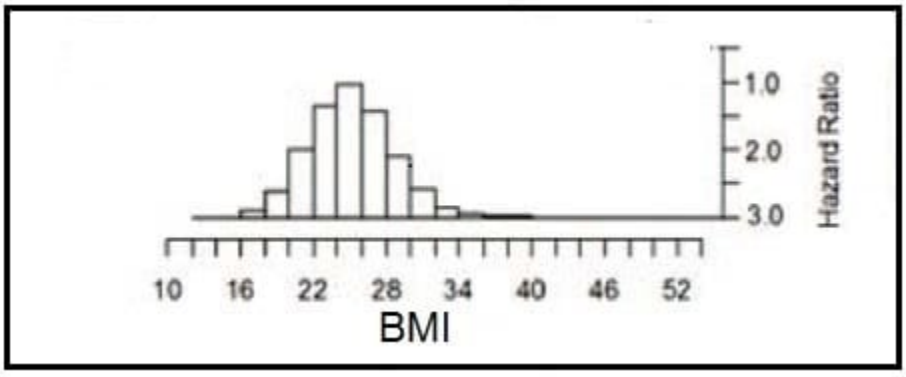


Figure 13.15: Relationship between BMI and the Hazard Ratio of mortality. Redrawn with data from Sun et.al. (2016).

Indeed, some experts have set lower cutoffs for overweight (i.e.,23–27.4kg/m^2) and obesity (> 27.5kg/m^2) for Asians based on their cardiovascular and type 2 diabetes risks (Al Kibria, 2019). However, WHO has not set a lower cutoff for Asians in view of the evidence that the risk appears to vary in different Asian populations. Instead, WHO recommends that a broad range of BMI categories should be used for reporting the prevalence of overweight in countries.

Several investigators have studied the determinants of both underweight and overweight in low- and middle-income countries in an effort to develop country-specific strategies to reduce or halt the double burden of malnutrition. (Al Kibria, 2019), for example, using data from the national Demographic and Health Survey (DHS) in Nepal, examined the prevalence and factors affecting underweight, overweight, and obesity among adults in 2016. They followed the WHO recommendation and estimated the prevalence by sex over a broad range of BMI categories which included three for underweight and seven for overweight, as shown in Table 13.7.

Table 13.7. Prevalence (with 95% confidence interval) of BMI categories among adult men and women, Nepal Demographic and Hea 2016. BMI: body mass index in kg/m2, IQR: Inter-quartile range. a Asian cutoffs for overweight and obesity start from these ranges, b World Health Organization cutoffs for overweight and obesity start from these ranges, respectively. Both cutoffs define underweigl 18.5kg/m2. From: (Al Kibria, 2019).

BMI cut-offs	Men	Women
≤ 16.00	2.1 (1.7–2.5)	3.5 (3.1–4.0)
16.00–16.99	2.9 (2.5–3.4)	3.9 (3.5–4.4)
17.00–18.49	10.1 (9.3–10.9)	10.5 (9.8–11.2)
18.50–22.99	49.8 (48.5–51.2)	43.2 (42.0–44.3)
23.00–24.99 [a]	15.9 (14.9–16.9)	14.1 (13.3–14.9)
25.00–27.49 [b]	11.5 (10.7–12.4)	11.6 (10.9–12.3)
27.50–29.99 [a]	5.2 (4.6–5.8)	7.7 (7.1–8.3)
30.00–32.49 [b]	1.9 (1.6–2.3)	3.4 (3.0–3.8)
32.50–34.99	0.1 (0.0–0.7)	1.5 (1.3–1.8)
35.00–37.49	0.1 (0.0–0.7)	0.5 (0.3–0.7)
37.50–39.99	0.1 (0.0–0.2)	0.1 (0.1–0.2)
≥ 40.00	0.1 (0.0–0.1)	0.1 (0.0–0.2)

In Nepalese adults, the overall prevalence of underweight (I.e., BMI < 18.5kg/m^2) was 16.7%, higher than the global prevalence, but comparable to the prevalence in South Asia (NCD Risk Factor Collaboration, 2017), whereas that of overweight (18.2%) and obesity (4.3%) was lower than the global average; 45.9% had a BMI within the normal range. Age, sex, and education level, household wealth status, place, ecological zone, and provinces of residence were all determinants of the extreme body weight categories (i.e., underweight, overweight, and obesity). Overall, higher education level and wealth status were positively associated with overweight / obesity, and inversely associated with underweight. Women had a higher prevalence of being both underweight and overweight / obese compared to males, consistent with earlier studies (Flegal, 2006).

Successive national DHS surveys have been used to track the rates of change in the prevalence of underweight and overweight in some countries. As an example, based on a meta-analysis of data on women of reproductive age from five successive national DHS surveys in Bangladesh from 1996 to 2011, the overall total prevalence of underweight decreased from 47% to 24% over this time period, whereas that of overweight increased from 4% to 16%. Moreover, this shift in BMI from underweight to overweight from 1996 to 2011 was most strongly associated with urban residence, age, higher socioeconomic status, and higher education attainment (Hoque et al., 2015). (Figure 13.16).

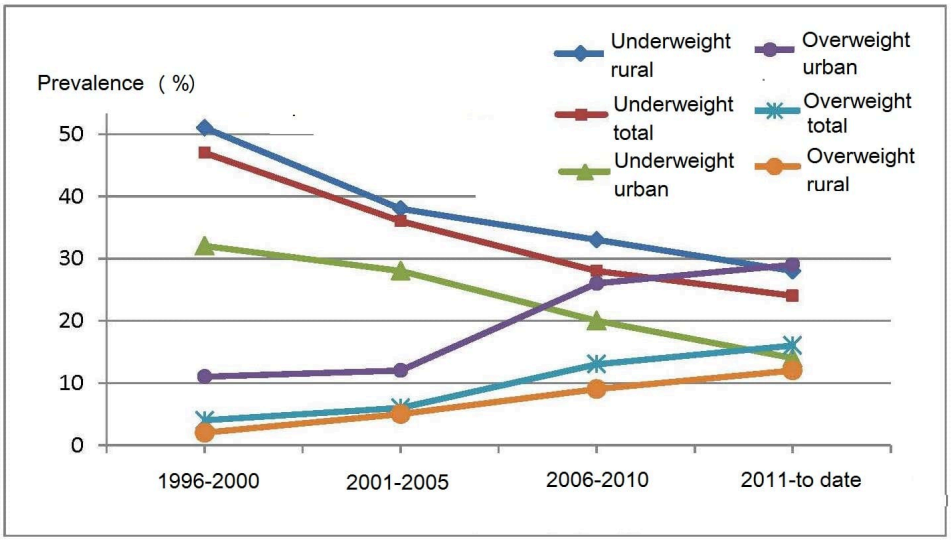


Figure 13.16. Prevalence of underweight and of overweight in Bangladeshi women. Redrawn from Hoque et al. (2015).

The increasing double burden of underweight and overweight in women during the preconception period is especially concerning as both are known to contribute to maternal and fetal complications during pregnancy (Salihu et al., 2009; Han et al., 2011). In a systematic review and meta-analysis of 34 studies, pre-pregnancy underweight (defined by a BMI < 18.5) contributed to a 32% higher risk of preterm birth, while for overweight women, the risk of preeclampsia and gestation diabetes mellitus approximately doubled, an effect that was even greater among women with pre-pregnancy obesity. These adverse effects of both underweight and overweight in the preconception period on pregnancy outcomes are likely to be amplified in adolescents or women with closely-spaced pregnancies and thus depleted nutrient reserves. These findings emphasize the importance of appropriate interventions to optimize maternal BMI in the pre-conception period (Dean et al., 2014).

13.3.8 Composite Index of Anthropometric Failure (CIAF)

The three conventional growth indices — weight-for-age, weight-for-height, and height-for-age — are described in Section 13.1; additional details are available in Chapter 10. Each index provides important information on different aspects of undernutrition (e.g., chronic versus acute) and can be used to select the most appropriate intervention.

In many low income countries, weight-for-age is the most commonly used growth index for children, with the proportion underweight (i.e., weight-for-age Z-score < –2) being used extensively globally as a proxy indicator to monitor progress in meeting the Millennium Development Goal to eradicate hunger. However, as discussed earlier and in detail in Chapter 10, underweight can be caused by stunting or wasting or a combination thereof and does not distinguish between them. Therefore, underweight when used on its own does not identify the sum of those children who are stunted and/or wasted, and thus provides an underestimate of the overall magnitude of undernutrition among children in the population. This is unfortunate because for policy and planning in a country, it is important to understand the overall scale of the problem of undernutrition to ensure sufficient resources can be allocated.

Consequently, an alternative classification system — termed the Composite Index of Anthropometric Failure (CIAF), has been developed by Nandy and Svedberg (2012), and is presented in Table 13.8. The original CIAF provided an aggregate measure to estimate the overall burden of undernutrition based on children whose weight and height were below the WHO Child Growth Standard.

Table 13.8. Categories of the composite index of anthropometric failure (CIAF). Modified from Savanur and Ghugre (2015).

Group	Description of the group	Definition
A	No anthropometric failure	Normal WAZ, HAZ, WHZ

Group	Description of the group	Definition
B	Wasting only	WHZ < –2 SD but normal WAZ and HAZ
C	Wasting and underweight	WHZ and WAZ < –2 SD but normal HAZ
D	Wasting, underweight & stunted	WHZ, WAZ and HAZ < –2 SD
E	Stunting and underweight	HAZ and WAZ < –2 SD normal WHZ
F	Stunting only	HAZ < –2 SD normal WAZ and WHZ
G	Stunted and overweight	HAZ < –2 SD and WHZ > +2 SD
H	Overweight only	WHZ > +2 SD and WAZ > +2 SD normal HAZ
Y	Underweight only	WAZ < –2 SD normal HAZ and WHZ

Comparison of the percentage of children classified as undernourished, based on the three conventional growth indices, and the CIAF using the new WHO Child Growth Standards for children under 5y have consistently shown more undernourished children using the CIAF than all the conventional indices (Nandy and Svedberg, 2012; Savanur and Ghugre, 2015). Children exhibiting multiple anthropometric failures are especially vulnerable.

Savanur, M. S., & Ghugre, P. S. (2015). Magnitude of undernutrition in children aged 2 to 4 years using CIAF and conventional indices in the slums of Mumbai city. Journal of health, population, and nutrition, 33, 3.

With the emergence of the dual burden of malnutrition, the original CIAF model of classification for under-nutrition is no longer sufficient. This has led to the extension of the CIAF model include two new groups: “stunted and overweight”; and “overweight only”, as shown in Table 13.8. In some studies, BMI-Z score has been used to define overweight instead of WHZ > +2 SD, with the BMI-Z score cutoff varying according to the age group of the children (Bejarano et al., 2019). The CIAF extended model provides policy-relevant information on the pattern of child malnutrition in any country.

As an example, the use of the extended CIAF (ECIAF) to classify malnutrition due to both undernutrition and over-weight in children (n=10,679) age 3–13.99y was explored in a study involving six Argentine provinces. The percentage of preschool children 3–4.99y with malnutrition based on ECIAF was lower than in the school children 5–13.99y (15.1% vs. 28.4%). In the whole sample, about 25% of the malnutrition was caused by undernutrition and 75% by over-weight (Bejarano et al., 2019).Kuwornu et al. (2020) also used the ECIAF to measure the overall burden of early childhood malnutrition in Ghana from 2008–2011 for children 6–59mos based on three national surveys.

This page titled 11.4: Screening systems to identify individuals at risk in a population (13.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

11.5: Applications of anthropometry in public health (13.4)

Box 13.5. Uses of anthropometry at the polulation level

- International comparisons of growth
- Identifying the determinants and consequences of malnutrition
- Targeting interventions
- Evaluating response to an intervention
- Nutritional Surveillance

The applications of anthropometry commonly used in public health are itemized in Box 13.5. Details of these applications are discussed briefly below. Recommendations for data collection, analysis and reporting of anthropometric indicators in children under-5y are available from WHO.

The reader is strongly recommended to follow these WHO guidelines to ensure harmonized methodology and reporting and thus facilitate comparisons of growth internationally. The WHO recommended age groups, (Table 13.9), Z-score cutoffs for abnormal growth, and the use of flags to identify implausible values should always be used. Details on how to identify and report implausible Z-score values are given in Section 13.1.2.

Table 13.9. Suggested age groups for use in the presentation of anthropometric data. From: WHO/UNICEF (2019).

Group	Age Classes
Newborn children, infants and pre-school children	0 to < 6mos 6 to < 12mos 12 to < 24mos 24 to < 36mos 36 to< 48mos 48 to < 69mos

Three methods have been recommended by WHO to evaluate cross-sectional anthropometric data for use in public health. These three methods are itemized in Box 13.6. To calculate the prevalence estimates for stunting and overweight based on national population data, the reader is advised to use the WHO tracking tool to set national targets and monitor progress. Note wasting is excluded from the tracking tool because of its high short-term variability. See WHO Global Target Tracking tool for more details. However, these methods can also be augmented with more complex statistical techniques to identify the determinants and consequences of malnutrition, and to set targets and/or assess the response to interventions, depending on the study objectives.

Box 13.6> Methods recomended by WHO to assess anthropometry

- Comparison of the frequency distribution of anthropometric indices with appropriate WHO growth reference data using Z-scores
- Summary statistics of the Z-scores: mean; median; SD; standard error (SE) with the 95% confidence interval (CI) for each growth indicator
- Calculation of number and proportion of individuals (as %) with anthropometric indices below or above a designated cutoff (95% CIs)

13.4.1 international comparisons of growth

(UNICEF/WHO/World Bank Group, 2017) highlight the levels and trends in child malnutrition across countries based on the thresholds of severity shown in (Table 13.10).

Table 13.10. Prevalence thresholds and corresponding labels for wasting, overweight and stunting in children under 5 years using the "novel approach". From: de Onis et al. (2019).

Wasting		overweight		Stunting	
Prevalence thresholds (%)	**Labels**	**Prevalence thresholds (%)**	**Labels**	**Prevalence thresholds (%)**	**Labels**
< 2·5	Very low	< 2·5	Very low	< 2·5	Very low
2·5–< 5	Low	2·5–< 5	Low	2·5–< 10	Low
5–< 10	Medium	5–< 10	Medium	10–< 20	Medium
10–< 15	High	10–< 15	High	20–< 30	High
≥ 15	Very high	≥ 15	Very high	≥ 30	Very high

As an example, Figure 13.17 depicts a modeled percentage of children under 5y affected by stunting in countries worldwide in 2020, categorized by the five prevalence thresholds for stunting presented in Table 13.10. Currently, no estimates exist on children suffering simultaneously from two forms of malnutrition such as stunting + overweight or stunting + wasting.

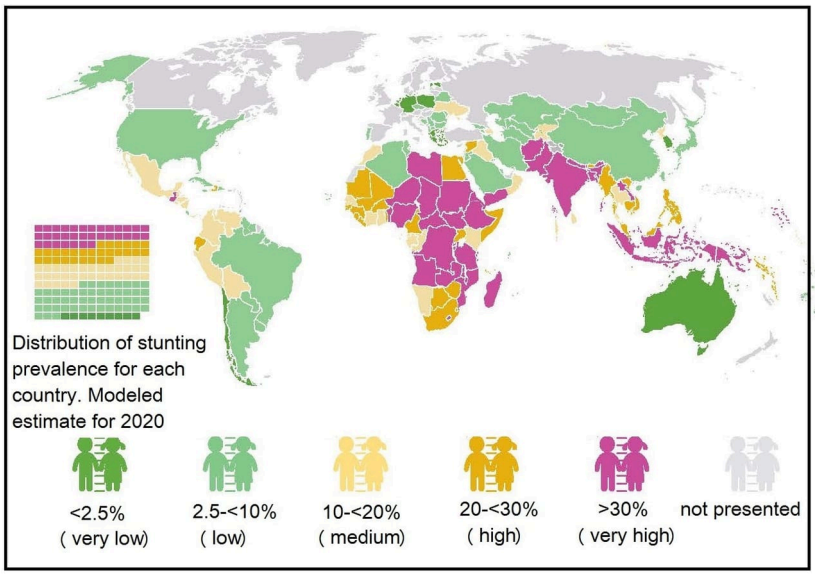


Figure 13.17. Modeled percentage of children < 5y affected by stunting in 2020. From: UNICEF, WHO, World Bank Group (2021).

As an alternative to only depicting a subset, the mean Z-score and standard deviation of the entire distribution of HAZ-scores can be calculated. If the mean Z-score is significantly lower than zero — the expected value for the reference distribution — then usually the entire distribution has shifted downward suggesting that most, if not all of the individuals, have been affected. This can be confirmed by plotting the entire distribution of the Z-scores against the reference distribution as shown in Figure 13.18.

Here all the children, with an overall mean Z-score of –1.83, were affected by some degree of linear growth retardation and an intervention is required for the entire community. This is termed a "population approach to targeting". In contrast, the stunted children, as defined by HAZ < –2 were only a subset of those with linear growth retardation.

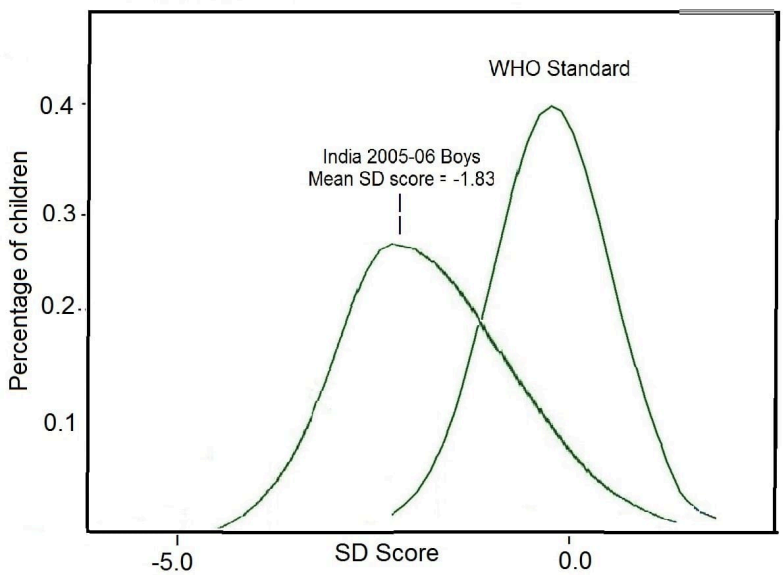


Figure 13.18. Distribution of length/height-for-age Z-scores of children from the Indian National Family Health Survey 2005–2006. Modified from de Onis and Branca (2016).

With the recognition that anomalies in the distribution of abdominal fat are as great a risk factor for disease as is excess body fat *per se,* waist circumference is increasingly used as a surrogate estimate of intra-abdominal fat content in adults, often in combination with BMI as a surrogate indicator of obesity. However, to date, the use of this combination in large scale population studies is limited.

Adult cutoffs for waist circumference that vary by sex and ethnicity have been developed (M: ≥ 102cm, F: ≥ 88cm, WHO, 2000) with lower cutoffs proposed for Asians (M: ≥ 90cm, F: ≥ 80cm, Lear et al., 2010) than Europeans, as reported for BMI. Nevertheless, more research is needed to establish whether specific cut-offs are needed for African-American, Hispanic, and Middle Eastern populations (Lear et al., 2010); see Chapter 11 for more details.

13.4.2 Identifying the determinants and consequences of malnutrition

The main direct determinants of malnutrition are frequently dietary intake and infection, both of which are affected by underlying determinants such as household socioeconomic conditions, food security, education, health services, or water, sanitation and hygiene (i.e., WASH). Anthropometric indices can be used in conjunction with household and WASH data, to examine factors associated with malnutrition as well as its consequences. Anthropometric indices are especially useful in extreme acute situations such as famine, as the determinants of very high levels of wasting are severe energy deficits. The latter can be confirmed by a rapid reduction in the prevalence of low weight-for-height following re-feeding.

In non-disaster situations, however, the causes of wasting, stunting, or overweight are much more difficult to identify because multiple factors may be involved, making it important to collect information or measure all potential risk factors. Aguayo et al. (2016) investigated the determinants of stunting and poor linear growth in children in the state of Maharashtra, India where 22.7% of children 0–23mos were stunted, of whom 7.4% were severely stunted. They used data from the 2012 comprehensive nutrition survey which included questionnaires focussed on the households, mothers, and infant and young child feeding (IYCF) practices, using the standardized IYCF indicators developed by WHO (2008).

In the state of Maharashtra, the mean HAZ was significantly lower (–1.74 vs. –0.25) and the prevalence of stunting significantly higher (40.5% vs. 9.2%) among children age 18–23mos than among children age 0–5mos. Multivariate regression analyses suggested that the most consistent determinants of stunting and poor linear growth in these Indian children < 23mos were birthweight and child feeding, women's nutrition and status, and household sanitation and poverty; for more specific details see Aguayo et al. (2016). Based on these results, policies, programs, and investments to improve linear growth and reduce the preva-

lence of stunting in the State of Maharashtra should be prioritized in four key areas, as shown in Box 13.7.

Box 13.7. Key areas for policy and program investment to improve linear growth and reduce stunting in the State of Maharashtra

- Improve women's nutrition and reduce low birthweight;
- Improve complementary foods and feeding practices for children age 6–23mos;
- Improve access to and use of sanitation facilities;
- Mitigate household poverty through effective social safety nets coupled with effective communication and counseling.

In a more recent study, factors associated not only with stunting but also with wasting and underweight, were each examined in children aged12–59mos using Demographic and Health Survey data (2007–2018) from 35 low- and middle-income countries (Li et al., 2020). The comparative importance of both direct and underlying factors associated with anthropometric deficits were investigated. In most countries, irrespective of the anthropometric indicator applied, household socio-economic status (assessed by household wealth and maternal education) and parental nutritional status (i.e., height and BMI) were the strongest factors associated with anthropometric undernutrition, consistent with findings elsewhere (Li et al., 2017).Nevertheless, several other factors associated with anthropometric undernutrition were also identified and were heterogeneous among countries. These findings highlight the importance of understanding the underlying mechanisms of child undernutrition within a specific country prior to formulating intervention policies and programs to reduce child undernutrition.

Overweight and obesity in childhood is also of increasing public health concern in both developed and developing countries, as noted earlier. Hence, identifying modifiable risk factors for child overweight and obesity that can be targeted through public health interventions is also important. In a report of Kenyan preschool children between 3–5y (n=1495) based on data from a nation-wide Demographic Health Survey (DHS), Gewa (2010) examined the associations between childhood overweight and obesity and selected maternal and child-related factors. Maternal overweight and obesity, higher levels of maternal education, being a large or very large child at birth, and being stunted were each associated with higher odds of overweight and obesity among these Kenyan preschoolers. Some of these factors are modifiable and could be targeted through appropriate public health interventions. For more discussion, see Gewa (2010).

Nevertheless, the interpretation of such cross-sectional data on the determinants of stunting and overweight is difficult. Anthropometry and the data from the households, mothers, and young children are collected at the same time during a cross-sectional survey, whereas both stunting and overweight reflect long-term cumulative influences which may have existed for a long time before the households were surveyed.

Child stunting, wasting, under-weight or overweight each have short- and long-term health consequences which are well documented (Black et al., 2008; Nahar et al., 2020). The capacity of anthropometric indicators to predict the risk of these adverse health consequences has been the subject of much research; it depends on several factors. These include the indicator selected, the specific risk under question, the age of the children, the baseline levels of the anthropometric deficit, and the health status in the study population at baseline. The findings vary according to the study setting. Nonetheless, information from these investigations is important because they can promote the survival of these “at risk” children through appropriate interventions.

Many settings in developing countries are characterized by poverty, infections, and inadequate dietary intakes. As a result, children may exhibit multiple anthropometric deficits simultaneously, which may amplify their risk for adverse health outcomes such as morbidity or mortality. To address this concern, the association between simultaneous anthropometric deficits and the risk of adverse health consequences have been studied by some investigators. As an example, Nandy and co-workers (2005) employed the composite index of anthropometric failure (CIAF, see Section 13.3.8) and a standard of living index (SLI) for each child to investigate the determinants and consequences of undernutrition among children < 3y in the 1998–1999Indian National Family Health Survey. The SLI was based on a range of 30 household assets and possessions and used as a proxy for household wealth and economic status. Table 13.11 shows that the undernourished children in each of the six subgroups were found to have a lower mean SLI score compared to children in the referent group (Group A). Note the subgroup of children with wasting, stunting and underweight (Group D) had the lowest mean SLI score, and those children with multiple anthropometric failures (i.e., those in subgroups C, E, and D) had lower mean SLI scores than those with single failures (Groups F, Y and B), or no failure (Group A).

Table 13.11. Subgroups of anthropometric failure among Indian children 0–3y (n=24,396), with scores of < 14: low standard of living; 15–24: medium standard; and scores25–67; high standard. Data from Nandy et al. (2005).

Group	% of Children	Mean SLI (95% CI)
A (no failure)	40.2	22.4 (22.2, 22.6)
F (stunting only)	10.1	20.1 (19.7, 20.5)
Y (underweight only)	5.9	19.4 (18.9, 19.9)
B (wasting only)	2.6	19.3 (18.6, 20.1)
C (wasting and underweight)	6.1	18.3 (17.8, 18.8)
E (stunting and underweight)	27.9	16.9 (16.7, 17.1)
D (wasting, stunting and underweight)	7.2	15.3 (14.9, 15.7)

Nandy et al. (2005) also collected information on child morbidity from maternal reports on the two weeks preceeding the survey, as recommended by WHO (1994; 1999). Three indicators of morbidity were constructed: diarrhea (i.e., reported in last two weeks); severe diarrhea (i.e., blood in stools); and acute respiratory infection (i.e., cough followed by breathing difficulties). Age-adjusted binary logistic regressions for anthropometric failure, diarrhea, severe diarrhea and acute respiratory infection (Table 13.12) showed children who were simultaneously wasted, stunted, and underweight (i.e., Group D) had the highest risk of both diarrhea and acute respiratory infection. These findings suggest that children with multiple anthropometric failures are at a greater risk of morbidity and more likely to come from poorer households, as reported elsewhere (Baig-Ansari et al., 2006). However, because this was a cross-sectional study with the morbidity cases recorded before the anthropometric measurements were taken, the possibility that the associations observed arose from reverse causality (i.e., restricted growth could be due to previous infection), cannot be ignored.

Table 13.12. Age-adjusted binary logistic regression analysis for anthropometric failure, diarrhea, severe diarrhea, and acute respirat 0–3y.Information on morbidity was provided by the child's mother or caregiver.

CI = confidence interval. [b] *p < 0.001* [c] *p < 0.05.*
Data from Nandy et al. (2005).

	Diarrhea	Severe diarrhea	Acu
Group	Odds ratio 95% CI	Odds ratio 95% CI	Odd
A (no failure)	1.00	1.00	1.00
B (wasting only)	1.06 (0.86–1.30)	0.45 (0.19–1.10)	0.92 (0.7
C (wasting and underweight)	1.45[b] (1.27–1.65)	1.19 (0.82–1.73)	1.28 (1.12
D (wasting, stunting and underweight)	1.72[b] (1.52–1.95)	1.95[b] (1.45–2.61)	1.39 (1.23
E (stunting and underweight)	1.54[b] (1.42–1.67)	2.03[b] (1.66–2.49)	1.34 (1.24
F (stunting only)	1.04 (0.92–1.17)	1.08 (0.78–1.48)	1.02 (0.9

Group	Diarrhea Odds ratio 95% CI	Severe diarrhea Odds ratio 95% CI	Acu Odd
Y (underweight only)	1.19[c] (1.03–1.37)	1.64 (1.17–2.29)	1.16[c] (1.01

The associations of child stunting, wasting, and underweight with increased risk of mortality are well recognized (Fawzi et al., 1997). In many cases, these studies have been cross-sectional, as noted earlier, with mortality estimates based on the effects of individual anthropometric indicators. However, as highlighted by Nandy et al. (2005), multiple anthropometric deficits may occur simultaneously, which may amplify the risk of mortality.

Khan and Das (2020) used the CIAF model (Section 13.3.8) to classify undernutrition and investigate risk of child mortality based on data on children < 5y from the National Family Health Survey in India in 2015–2016. Although no causal inferences could be drawn, their results indicated that districts with a higher burden of multiple anthropometric failures based on the CIAF index were associated with an elevated risk of infant and child mortality; see Khan and Das (2020) for more discussion of the findings.

Prospective studies, unlike cross-sectional studies, provide the best information about causation of disease and the most direct measurement of the risk of developing disease. McDonald et al. (2013) were the first to estimate the mortality risk associated with multiple anthropometric deficits in prospective cohorts based on a meta-analysis of data from 10 prospective studies in Africa, Asia, and Latin America. Unfortunately, cause-specific mortality was not investigated due to insufficient statistical power. In this study, the recent anthropometric status of the children was classified into seven subgroups shown in Table 13.13.

Table 13.13. Baseline distribution of anthropometric status based on data from 10 prospective studies in Africa, Asia, and Latin America. From: McDonald et al. (2013).

Classification of anthropometric status	*n* (%)
Not stunted, wasted, or underweight	27,794 (51.7)
Stunted only	7475 (13.9)
Wasted only	1025 (1.9)
Underweight only	1358 (2.5)
Stunted and underweight but not wasted	11,501 (21.4)
Wasted and underweight but not stunted	1655 (3.1)
Wasted, stunted, and underweight	2953 (5.5)

Mean age of the children at baseline ranged from 5wk to 2.8y and the mean follow-up time ranged from ≈ 35wk to 85wk. A significantly elevated risk of all-cause mortality was reported among children with 1, 2, and 3 anthropometric deficits compared with children with no deficits, with a dose response relationship in relation to the number of deficits. Moreover, for children wasted and underweight but not stunted, risk of mortality was greater than for those who were stunted and underweight but not wasted; see McDonald et al. (2013) for further details.

More studies are needed to establish whether there are unique determinants of multiple anthropometric deficits, as well as to identify the combination of anthropometric deficits associated with the greatest risk of mortality. With this knowledge, interventions could be targeted to the subgroup of children who may benefit the most.

13.4.3 Targeting interventions

Anthropometry can be used as a screening tool to identify the areas of greatest need (e.g., areas with a high prevalence of stunting or wasting or multiple anthropometric deficits) and, hence, likely to gain the most benefit from any intervention. WHO and UNICEF have classified the severity of malnutrition in young children < 60mos into five categories based on the percentage prevalence threshold of wasting, overweight, and stunting(de Onis et al., 2018) (Table 13.14).

Table 13.14. Prevalence thresholds, corresponding labels, and the number of countries (n) in different prevalence threshold categorie children under 5 years using the "novel approach". From: de Onis et al. (2018).

Wasting			overweight			Stunting
Prevalence thresholds (%)	Labels	(n)	Prevalence thresholds (%)	Labels	(n)	Prevalence thresholds (%)
< 2·5	Very low	36	< 2·5	Very low	18	< 2·5
2·5 – < 5	Low	33	2·5 – < 5	Low	33	2·5 – < 10
5 – < 10	Medium	39	5 – < 10	Medium	50	10 – < 20
10 – < 15	High	14	10 – < 15	High	18	20 – < 30
≥ 15	Very high	10	≥ 15	Very high	9	≥ 30

The fifth threshold labeled "very low" and of no public health concern was included across all three indicators to reflect the expected prevalence of 2.3% (rounded to 2.5%) below/above 2 SDs from the median of the WHO Child Growth Standard. The higher four prevalence thresholds are defined by multipliers of this "very low" level rounded to 2.5%. Note that for wasting and overweight, the thresholds are the same: "very low" (< 2.5%), "low" (≈ 1–2 times 2.5%); and "very high" (> ≈ 6 times 2.5%). For stunting, the new thresholds are "very low" (< 2.5%); "low "(≈ 1–4 times 2.5%); "medium" (≈ 4-8 times 2.5%); "high" (≈ 8–12 times 2.5%) and "very high"(≈ 12 times 2.5%).

The number of countries in the different threshold categories for wasting, overweight, and stunting, shown in Table 13.14, is based on data accessed in 2018 from the WHO Global Database on Child Growth and Malnutrition and drawn from nationally representative surveys in 134 countries of children < 5y.

Comparison of the prevalence estimates for each anthropometric indicator can trigger countries to identify the most appropriate intervention program to achieve "low" or "very low" prevalence threshold levels. For example, in Table 13.14, in the countries in which the prevalence of stunting was "very high" (i.e., n=44), intervention activities should focus on resolving factors associated with stunting. These often include increasing the availability of food, and improving diet quality, hygiene, and potable water supplies, as well as programs to prevent and treat infectious diseases (Gorstein et al., 1994).

13.4.4 Evaluating response to an intervention

When using anthropometry to evaluate the performance of an intervention, at least two measurements, taken before and after the intervention period, should be recorded, where feasible. In child growth monitoring and promotion programs, for example, serial measurements of weight, and sometimes length/height, are taken on children attending child welfare clinics or community health centres. These measurements are then plotted manually on distance growth charts, preferably those developed by WHO (Figure 13.3), to provide information on the pattern of growth over time and the response to the intervention, where appropriate; see Section 13.2.2 and Section 13.2.3 for more details. When choosing an anthropometric indicator to monitor the response, the time lag between the start of the intervention and the time when a response can be anticipated must always be considered. As noted in Section 13.2.3, children who are wasted, respond to a nutrition intervention by gaining weight first, only later catching up in linear growth. Hence, plotting weight-for-length/height (if available) or weight-for-age on the WHO Child Growth Standard distance chart would be appropriate.

For large scale public health programs, the purpose of an evaluation is to influence decisions about the effectiveness of the programs. These decisions may include to continue, expand, modify, strengthen, or discontinue a program. Habicht et al. (1999) have reviewed the various designs recommended for evaluating public health programs. The designs vary depending on the purpose of the evaluation and the level of precision required and are summarized in Box 13.8.

Box 13.8. Types of intervention designs recommended to achieve various levels of inference, focusing on impact evaluation

- **Adequacy — Purpose of the evaluation:** determine whether the prevalence of stunting is adequate (according to a predetermined criteria (e.g., very low prevalence threshold: < 2.5%).

- **Types of design:** Cross-sectional survey of the prevalence of stunting (i.e., HAZ < –2 Z-score) at a certain point in time, and comparison of prevalence findings with the pre-established criteria of adequacy (i.e., < 2.5% stunting).

- **Plausibility — Purpose of the evaluation:** determine whether it is plausible to conclude that the intervention contributed to current prevalence of stunting or to the observed changes in the prevalence of stunting.
  - **Types of design:** Quasi-experimental or:
    - Cross-sectional survey with treatment and comparable control group;
    - Longitudinal study with before and after measurements (examine changes in treatment group only);
    - Case-control study with measurement of cases compared with matched comparable controls.
- **Probability — Purpose of the evaluation:** determine with a pre-established level of probability that the impact on stunting prevalence is due to the intervention.
  - **Types of design:** Double-blind, randomized experimental design:
    - Randomization of the intervention;
    - Before and after measurements with intervention and control groups;
    - Both subjects and researchers are blind to the treatment during the intervention and the analyses.

Modified from Habicht et al. (1999) and International Zinc Consultative Group (IZiNCG) (2004).

Adequacy evaluation is undertaken when it is not feasible to include a comparison or control group. Instead, the intervention is evaluated on the basis of a cross-sectional survey in which the prevalence of abnormal growth at a certain time-point is compared with the prevalence of abnormal growth defined by a pre-established criterium of adequacy. This could include, for example, the "very low" prevalence threshold for abnormal growth shown for children under 5y in Table 13.10. Adequacy of the intervention would be achieved if the evaluation showed that the prevalence of stunting, for example, at the time of the evaluation, was less than the pre-determined "very low" prevalence threshold of 2.5%.

In an adequacy evaluation, program activities cannot be causally linked to the observed changes. Nevertheless, the adequacy evaluation may provide the necessary reassurance that the expected goals of the program are being met, thus ensuring the continued support for the program. Examples of public health programs in which adequacy evaluation has been applied include a 10-year Micronutrient and Health (MICAH) program implemented from 1996–2005 in selected areas of Ethiopia, Ghana, Malawi, and Tanzania by World Vision (Berti et al., 2010), and the large-scale program to improve infant and young child feeding (IYCF) and nutrition in Ethiopia by the Alive and Thrive initiative (Menon et al., 2013).

Plausibility evaluation is conducted when decision makers require a greater degree of confidence that the observed changes were indeed due to the program. For an evaluation to be plausible, it must be able to show that the intervention appears to have an effect above and beyond other confounding factors which might have caused the observed effects. Hence, although this design does not require randomization of intervention and control activities to the comparison groups, nevertheless, attempts must be made to control for the influence of possible confounding factors. This can be achieved at the intervention design stage by the careful selection and matching of the comparison or control groups before the evaluation is begun. Alternatively, the influence of confounders or modifiers can be assessed afterwards at the analysis stage, when stratification or multivariate analyses can be used, provided information on potential confounders or modifiers has been collected or measured (Habicht et al., 1999). Of the plausibility designs summarized in Box 13.8, those of choice are the quasi-experimental or case-control designs. For details on selecting appropriate control or comparison groups for plausibility evaluation, see Rogers (2014) and White & Sabarwal (2014). For guidance on the statistical strategies to control for confounders and modifiers, the reader is advised to consult a biostatistician.

Probability evaluation designs provide the highest level of confidence that the intervention caused the outcome, and are considered the gold standard method. They require the use of a randomized, controlled, double-blind design which aims to ensure there is only a small known probability (usually < 0.05) that the observed difference in the prevalence between the intervention and control communities (or individuals) is due to confounding, bias, or to chance. With the double-blind randomization of the subjects to intervention and control groups, baseline differences between the intervention and control groups are minimal.

Figure 13.19. provides an example of the method used in a probability evaluation design of Alive and Thrive. This design is especially appropriate when evaluating long-running programs, such as 6y for the Alive and Thrive initiative. Note in particular that the intervention and control groups in this design are similar at baseline as a result of the randomization. Over such a long time frame, non-intervention-related changes in underlying determinants of nutrition, such as income, food security, education, health services, or water and sanitation may occur. These may generate changes in the control group during the duration of the intervention as shown in Figure 13.19, highlighting the importance of a control group, without which the true impact of the inter-

vention would have been overestimated.

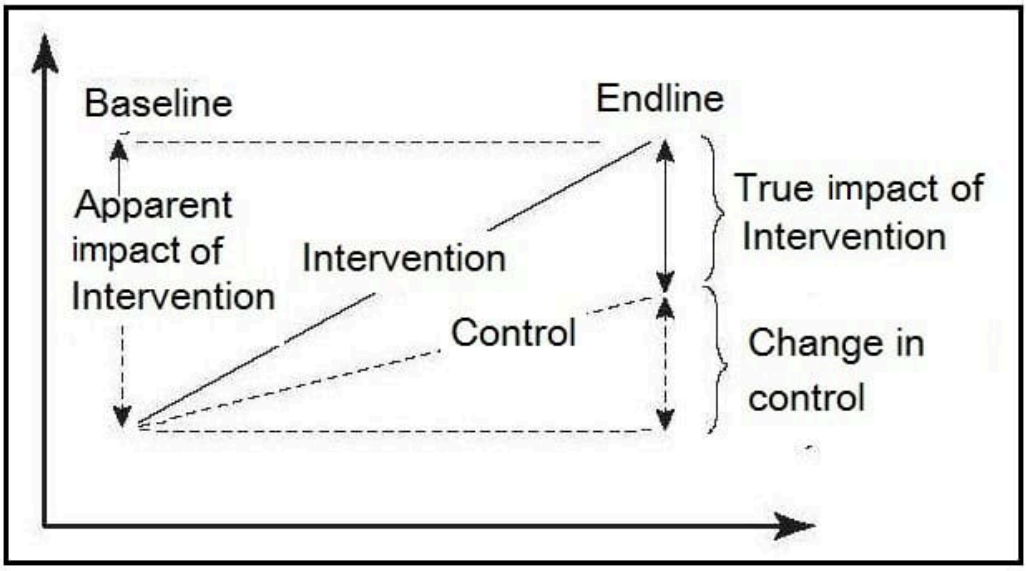


Figure 13.19. Illustration of the double-difference approach (i.e., difference between the changes in outcomes for intervention and control groups in the baseline and endline surveys) with randomization. This approach allows a "clean" estimate of the impact attributable to the program by taking into account the changes in outcomes that have occurred in the control group as a result of nonprogram factors.

In order to attribute the changes in the outcomes to the interventions implemented, Alive and Thrive used the difference-in-difference (DID) or double difference method. This method compares the difference in changes in outcomes over time between the intervention and comparison groups to estimate impact, and is said to yield a reliable attribution of impact to an intervention. For more details see White and Sabarwal (2014).

Challenges may arise however, when interpreting the data of the impact of the intervention when nonintervention-related changes occur, and additional statistical techniques may be required to understand and inform policy makers about which underlying factors are most important in explaining the improvements in linear growth; see Nguyen et al. (2017) for more details. Gertler et al. (2011) also provide details on statistical methods to evaluate large-scale programs.

Table 13.15 shows an example of a probability evaluation based on a cluster-randomized controlled trial that examined the effectiveness of a nutrition education intervention program in Peru. The cohort of young children, recruited at 6mos, were from a poor periurban area. The intervention was delivered through the health services with the aim to enhance the quality and coverage of nutrition education programs. The primary outcome was growth, and the impact of the intervention program on anthropometric indicators was assessed using the "difference-in-difference" (DiD) approach and is shown in Table 13.15. Note that the growth differences were adjusted for socioeconomic status, hygiene score, and birthweight, after accounting for cluster randomisation.

The adjusted mean changes in weight gain, length gain, and Z-scores at 18mos were all significantly greater for the children in the intervention area than for those in the control area. In addition, children in the intervention group were less likely to be stunted at 18mos that those in the control group; see Penny et al. (2005) for more details.

*Table 13.15. Growth at 18mos by intervention group in Peruvian children. Differences adjusted for socioeconomic status, hygiene score, and birthweight. *Odds ratio (95%CI) **Mean of the regression gradient for length-for-age. From: Penny et al. (2005).*

	intervention	Control	Adjusted differences*	P
Mean weight at 18mos (kg)	10.77 (1.16)	10.48 (1.02)	0.199 (-0.033 to 0.431)	0.093
Mean length at 18mos (cm)	79.36 (2.74)	78.29 (2.66)	0.714 (0.146 to 1.282)	0.014

	intervention	Control	Adjusted differences*	P
Mean WAZ at 18mos	-0.33 (0.90)	-0.62 (0.83)	0.194 (0.008 to 0.380)	0.041
Mean LAZ at 18mos	-0.81 (0.80)	-1.19 (0.83)	0.272 (0.099 to 0.445)	0.002
Mean WLZ at 18mos	0.15 (0.87)	0.05 (0.79)	0.048 (-0.139 to 0.237)	0.609
Linear growth velocity **	-0.031 (0.043)	-0.057 (0.050)	0.021 (0.012 to 0.031)	<0.0001
Number stunted at 18mos	8 of 171 (4.7%)	26 of 165 (15.8%)	3.035 * (1.207 to 7.636)	0.018

Abnormal changes in the rate of growth can be identified much earlier when growth velocity instead of distance growth is assessed, as noted previously (Section 13.2.2). However, growth velocity can only be assessed in interventions when the same children have been measured on more than one occasion, as was performed in the study of Penny et al. (2005). In this study in Peru, length velocity was very significantly greater in the intervention group than in the controls at 18mos.

For short-term interventions, assessment of weight gain or weight velocity rather than length or height velocity is more appropriate and easier to measure. As an example, in a randomized controlled trial, Sie et al. (2018) evaluated the effect of a 5-day course of three different antibiotics (amoxicillin, azithromycin, and cotrimoxazole) on growth among preschool children in Burkina Faso. Short-term growth was assessed from weight and height measurements taken at baseline and 30 days following the last medication dose from which weight velocity (g/kg/d), change in weight (g) and height (in cm), together with WHZ, WAZ, and HAZ- scores, were calculated. Results confirmed that children randomized to amoxicillin gained significantly more weight, had higher weight velocity and WHZ and WAZ-scores compared to their placebo siblings, but there were no differences in weight gain in those children randomised to azithromycin or cotrimoxazole. Moreover, as expected, no difference in height gain or HAZ were detected over the 30day study period across any of the study arms. These findings highlighted that amoxicillin may have short-term growth-promoting effects in children in settings with a high prevalence of undernutrition and child mortality.

13.4.5 Nutritional surveillance

The overall purpose of a nutritional surveillance program is to gather, interpret, and disseminate information about nutrition. WHO (1995) defines nutritional surveillance as:

> The continuous monitoring of the physical status of a population, based on repeated surveys or on data from child health or growth-monitoring programs.

The nutrition information collected can be used to achieve the objectives shown in Box 13.9.

Box 13.9. Objectives of nutrition surveillance

- Aid long-term planning in health and development;
- Provide input for program management and evaluation;
- Give timely warning of the need for intervention to prevent critical deteriorations in food consumption.

To achieve these objectives, the nutrition information collected must be:

- population-based;
- decision and action orientated;
- sensitive and accurate;
- relevant and timely;
- readily accessible;

- communicated effectively.

Modified from Jerome and Ricci (1997).

The U.S. has conducted a comprehensive national surveillance system since 1959 on the health status of US residents as part of the National Health and Nutrition Examination Surveys (NHANES). Data on anthropometry, demographic and socioeconomic status, dietary and health-related measures are collected.

The U.K. also has a surveillance system designed to assess the food consumption, nutrient intakes and nutritional status of the general population 18mos and older living in private households. The system is termed the National Diet and Nutrition Survey Rolling Program and began in 2008.

Successive national DHS surveys have also been used to track the rates of change in the prevalence of stunting, underweight, and overweight in some countries. As noted earlier (Section 13.3.7), in a meta-analysis of data on women of reproductive age from five successive national DHS surveys in Bangladesh from 1996 to 2011, the prevalence of underweight was reported to decrease from 47% to 24% over this time period, whereas that of overweight increased from 4% to 16% (Figure 13.16). This shift in BMI from underweight to overweight from 1996 to 2011 was most strongly and positively associated with urban residence, age, higher socioeconomic status, and higher education attainment (Hoque et al., 2015).

Recently, WHO has provided countries with tools to develop or strengthen their surveillance systems so they have the national capacity to monitor changes in the global nutrition target indicators shown in Box 13.2. These indicators focus specifically on childhood stunting, anemia in women of reproductive age, low birthweight, overweight, and exclusive breastfeeding.

The tools developed by WHO/UNICEF include the provision of six policy briefs, each covering one of the global targets with details of the actions that should be prioritized to meet the global target. In addition, a web-based tracking tool has also been developed to assist countries to set national targets and chart their progress of achieving the six global goals. Details of the indicators recommended to monitor each target are available from WHO (2013). More details of these tools are available from WHO (2019).

The rising global burden of chronic, non-communicable diseases (NCDs) has also led WHO to adopt a Global Monitoring Framework that includes 25 key indicators to track progress in the prevention and control of NCDs. This initiative prompted WHO in 2002 to develop a stepwise approach to NCD surveillance (STEPS) (WHO STEPS Surveillance Manual, 2005). The approach involves a standardized but flexible framework for countries to monitor the main NCD risk factors based on national household surveys of adults aged 18 to 69y, preferably repeated every 3 to 5 years.

In step 1 of this approach, self-reported information on demographics and behavioral risk factors are collected through in-home interviewer-administered questionnaires. This is followed by step 2 in which physical measurements of height and weight (to calculate BMI), waist circumference, and blood pressure are undertaken. Step 3 is performed at a local clinic or health center for biochemical assessment for blood glucose, blood lipids, and urinary sodium (Riley et al., 2016).

Since 2009, WHO has introduced eSTEPS for the data collection which has led to a more streamlined and standardized approach to quality control. WHO also provides technical support through training workshops for the implementation of STEPS, as well as STEPS “Data to Action” workshops to strengthen the in-country development of national NCD policies and programs. Ng et al (2014) provided a systematic analysis of the global, regional, and national prevalence of overweight and obesity in children and adults from 1980–2013. Globally, the proportion of children, adolescents, and adults with a BMI indicative of overweight and obesity has increased over this time period in both developed and developing countries, although for adult obesity the increase has stabilized since 2006. Moreover, age and sex-related differences in patterns of adult overweight and obesity exist, with men having higher rates than women in developed countries in 2013, peaking at ≈ 55y, whereas in developing countries, rates are higher in women than men, peaking at ≈ 45y (Figure 13.20). Nevertheless, the levels of overweight and obesity for both men and women in developing countries continues to be lower than that of developed countries.

Data generated through STEPS has been used in a number of projects.

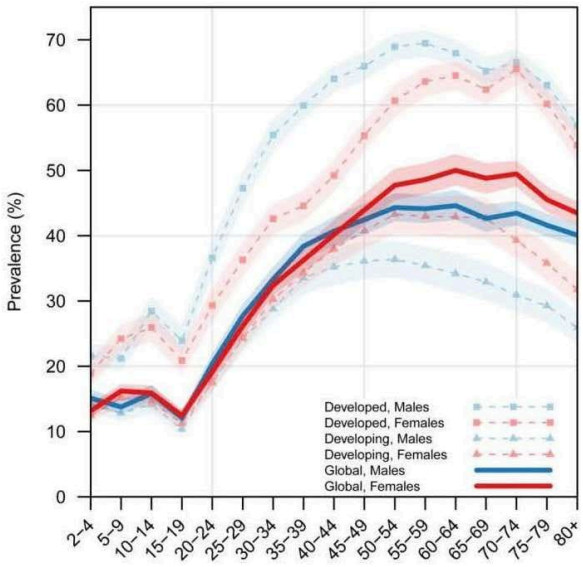


Figure 13.20: Prevalence of overweight and obesity (BMI ≥ 25) by age and sex, 2013. Modified from: Ng et al. (2014).

There has been a concerted effort over the past decade to strengthen nutritional surveillance globally. This effort is not only good public health practice, but it has also increased public and government awareness of the extent of the problems of malnutrition in both developing and developed countries. Future developments in surveillance may include methods that quantify intra-abdominal and subcutaneous adipose tissue, and measure physical activity patterns using wearable technology, and assay hemoglobin A1c as an indicator of diabetes. For more discussion of the challenges, limitations, and future directions of STEPS, see Riley et al. (2016).

As noted in Section 13.1.3, questions have been raised about the validity of interpreting changes in HAZ scores across time as representative of growth faltering or catch-up growth. Several studies have reported that HAZ scores increased despite increasing height deficits, highlighting that that the use HAZ scores rather than absolute height differences from the median (HADs) may result in opposite conclusions about the occurrence of catch-up growth (Lundeen et al., 2014). Hence caution is required when defining catch-up growth. The authors recommend that absolute height differences should always be included along with HAZ-scores when evaluating catch-up growth in height in populations with a high prevalence of growth failure. More research is needed to determine which of the two metrics is better at predicting different outcomes.

This page titled 11.5: Applications of anthropometry in public health (13.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12: BODY COMPOSITION- LABORATORY METHODS (CHAPTER 14)

Abstract

Several indirect in vivo laboratory methods are available to assess body composition. Their selection depends on the study objective, the precision and accuracy required, the cost, convenience to the subject and their health, and equipment and technical expertise available. This chapter describes both non-scanning and scanning in vivo laboratory techniques. Non-scanning techniques include total body potassium, total body water (via isotope dilution or bioelectrical impedance analysis (BIA)), neutron activation analysis, densitometry (via under-water weighing or air-displacement plethysmography), and total body electrical conductivity. Four scanning techniques are included: computerized tomography, magnetic resonance imaging, dual energy X-ray absorptiometry (DXA), and ultrasound. Each technique generates body composition data in a different way, so the results are not interchangeable. For each technique, the characteristics, including both the advantages, limitations and assumptions applied to generate the body composition data, are outlined. In addition, their potential applications are summarized, with emphasis on the alterations in the relative proportions of the body components that may occur in certain life-stage groups or disease states. Such alterations may invalidate the determination of fat and fat-free mass in a 2-component model of body composition. In a 2-component model, constants for the hydration and density of fat-free mass which ignore the inter-individual variability in these properties, are often assumed.

Many factors have the potential to alter values for these assumed constants, especially growth and maturation in children, aging, pregnancy, and obesity. Consequently, researchers have developed specific constants for the hydration and density of fat-free mass specific for these circumstances. Their use will improve the estimates of fat-free mass when applying the simple 2-component model based on measurements of total body potassium, total body water, or whole-body density. From a comparison of nine in vivo body composition methods conducted by Field and co-workers (2015), measurement of densitometry via air displacement plethysmography was selected as the method with the highest degree of accuracy and reliability and with the least degree of technical error to track and monitor whole-body composition across the lifespan, provided any alterations in the relative proportions of the body components, are taken into account.

In clinical patients with certain disease states, over- or underhydration and abnormalities in mineral mass may occur, resulting in substantial variability in both the hydration and density of fat-free mass. In these circumstances, application of multi-component models that include measurements of protein and/or bone minerals (via neutron activation or DXA) as well as total body water and density will minimize assumptions related to the structure, hydration, and density of fat-free mass. The 4-component model is considered the criterion method whereby whole-body composition can be most accurately assessed. A simplified version based on measurements from both DXA and BIA holds promise for monitoring conditions in certain disease states.

The four scanning techniques are widely used in clinical settings to quantify in vivo body composition, especially at the tissue-organ level, when investigations of bone density, skeletal muscle mass, and the deposition of visceral ectopic fat with their corresponding relationship with osteoporosis, sarcopenia, or cardiometabolic risk, respectively, are needed.

12.1: An introduction to techniques used to measure body composition (14.0)

12.2: Chemical analysis of cadavers (14.1)

12.3: Total body potassium (TBK) (14.2)

12.4: Total body water from isotope dilution (14.3)

12.5: Multiple dilution methods (14.4)

12.6: In vivo activation analysis (14.5)

12.7: Densitometry (14.6)

12.8: Total body electrical conductivity (14.7)

12.9: Bioelectrical impedance (14.8)
12.10: Computerized tomography (14.9)
12.11: Magnetic resonance imaging (14.10)
12.12: Dual energy X-ray absorptiometry (14.11)
12.13: Ultrasound (14.12)
12.14: Summary - Method comparisons (14.13)

This page titled 12: Body Composition- Laboratory Methods (Chapter 14) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.1: An introduction to techniques used to measure body composition (14.0)

Accurate methods for measuring body composition are required in investigations of obesity, malnutrition, weight loss following bariatric surgery, muscle wasting, sarcopenia, osteopenia, and osteoporosis. Body composition information is also used to establish the appropriate prognosis and treatment of hospital patients, and with longitudinal assessment, to monitor the effects of interventions on body composition (Lemos & Gallagher, 2017).

Selection of the method to measure body composition depends on the required precision and accuracy, the study objective, cost, convenience to the subject, their health, and equipment and technical expertise available(Lukaski, 1987). Methods based on multi-component models that include analysis of protein and minerals, minimize assumptions related to tissue density, hydration, and structure. This is important because in malnourished individuals, the elderly, and subjects with metabolic disturbances, the relative proportions of the body components is often altered, and losses of protein, fat, and bone mineral content may occur, often in association with the rapid accumulation of water. Changes such as these invalidate the determination of fat and the fat-free mass in the 2-component model.

Absolute validity cannot be assessed for any of the indirect in vivo body composition methods because the gold standard for body composition analyses is cadaver analysis. Instead, only relative validity can be assessed, defined as the comparison for each subject of the results from the “test” method with the results from another method, termed the “reference or criterion” method; the latter having a greater degree of demonstrated validity. A 4-component model is now considered sufficiently accurate to act as a reference or criterion method, but its use in many settings is limited because of the expensive and sophisticated technology required. Multiple statistical approaches can be used to establish the validity of the “test” method compared with a reference method. They include regression and correlation analyses, paired t tests, and more recently, Bland-Altman analysis; see Earthman (2015) for further details.

The characteristics of the various procedures used for measuring body composition are summarized in Box 14.1. This list includes both non-scanning and scanning techniques. Detailed sections (14.2-14.12) describe each of these indirect in vivo methods now available to assess body composition. Comments on these methods are also given, along with the assumptions used, and the advantages and disadvantages of each method. Scanning techniques such as computer tomography (Section 14.9), magnetic resonance imaging (Section 14.10) and whole body dual energy X-ray absorptiometry (DXA) (Section 14.11) are included together with a discussion of their clinical importance (Lee et al., 2019; Neeland et al., 2019). Each technique generates body composition data in different ways, so the methods are not interchangeable. Methods with the lowest cost are often the most imprecise.

Methods employing the 2-component model (i.e., body fat and fat-free mass) include total body potassium, total body water via isotope dilution or bioelectrical impedance, densitometry via hydrostatic weighing or air-displacement plethysmography, and total body electrical conductivity. Such methods are not suitable for clinical populations when the basic assumptions of the 2-component model are often invalid. Instead, in these populations, techniques using a 3, 4, or 5-component model should be applied. For example, (Section 14.11) has the capacity to generate data that can be used with a 3-component model. See Lohman (1986) and Pietrobelli et al. (1996) for further details.

Three scanning techniques — computerized tomography (Section 14.9), magnetic resonance imaging (Section 14.10), and DXA (Section 14.11) — can be used to quantify components (e.g., skeletal muscle, bone, visceral ectopic fat) at the tissue-organ level of body composition as well as to assess the relative proportions of the fat-free mass, body fat, and bone mineral content. Of these, only DXA has been recommended for the assessment of fat mass in patients with a variety of disease states; the use of DXA for the assessment of fat-free mass is not recommended for clinical populations because its validity for assessment of fat-free mass in any clinical population remains unknown (Sheean et al., 2020).

Box 14.1. Scanning and non-scanning laboratory techniques used to measure body composition.

- **14.1 Chemical analysis of cadavers**
  Cadaver analysis provides the gold standard data on body composition, but the use of such data is limited by ethical barriers. Results from a few older studies (1945-1968) are mostly based on adults who had died because of illness.
- **14.2 Total body potassium (TBK)**
  TBK is measured by counting radiation from naturally occurring ^{40}K in a whole body counter. Required equipment is only found in specialized facilities. Estimates of the fat-free mass can be derived from the TBK.
- **14.3 Total body water from isotope dilution (TBW)**
  A tracer dose of water, usually labeled with the stable isotope ^{2}H, is given orally or intravenously and then allowed to equilibrate. The concentration of the isotope in serum, urine, or saliva is measured and TBW calculated from dilution observed following equilibration. Obesity, pregnancy, and wasting disease increase TBW.
- **14.4 Multiple dilution methods**
  Typically, multiple dilution involves determining both total body water via isotope dilution and extracellular water (ECW), the latter using a tracer such as bromide that does not enter the intracellular space. The difference between these two measurements (i.e., TBW – ECW) reflects the intracellular water.
- **14.5 In vivo activation analysis**
  Radioactive isotopes of N, P, Na, Cl, Ca are created by irradiating the subject. The resulting γ-radiation is measured using a whole body counter. Subjects are exposed to radioactivity. Sensitivity varies with the element. Required equipment is only found in specialized facilities.
- **14.6 Densitometry**
  Body density is derived from measurements of body mass and body volume. The latter is calculated from: (a) the apparent loss of weight when the body is totally submerged in water — difficult with young children, the elderly or sick patients — or, (b) air-displacement or water-displacement plethysmography.
- **14.7 Total body electrical conductivity (TOBEC)**
  Subject lies supine in a solenoid coil through which a 5MHz current is passed. The conductivity value of the subject is obtained by subtracting the background value when the coil is empty. Edema, ascites, dehydration, electrolyte balance and variations in bone mass all interfere with the conductivity reading.
- **14.8 Bioelectrical impedance (BIA)**
  The impedance to a weak electrical current passed between the right ankle and right wrist of a subject in supine position is measured. Edema, ascites, and dehydration invalidate single frequency measurements. Multifrequency measurements allow estimation of both total and extracellular compartments.
- **14.9 Computerized tomography (CT)**
  Method measures attenuation of X-rays as they pass through tissues, the degree of attenuation being related to differences in physical density of the tissues. An image is reconstructed from the matrix of picture elements. Exposure to ionizing radiation limits use of CT for pregnant women or children. Expensive equipment.
- **14.10 Magnetic resonance imaging (MRI)**
  Imaging involves placing a subject in a very strong magnetic field and observing the relative differences in behavior of ^{1}H protons in lean and adipose tissue. No exposure to ionizing radiation but equipment bulky and expensive.
- **14.11 Dual energy X-ray absorptiometry (DXA)**
  Utilizes the attenuation of a dual energy X-ray beam, often during whole-body scanning. New fan-beam technologies replacing earlier pencil-beam techniques lead to lower X-ray doses and improved spatial resolution. High precision method, but results are calibration dependent and differences between various equipment manufacturers can be significant.
- **14.12 Ultrasound**
  High-frequency sound waves from a combined ultrasound source and meter pass through adipose tissue to the adipose-muscle tissue interface. At the interface, some sound waves are reflected back as echoes, which are translated into depth readings via a transducer. CT, MRI, or DXA provides higher degree of structure resolution than ultrasound.

This page titled 12.1: An introduction to techniques used to measure body composition (14.0) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Chemical analysis of cadavers (14.1)

Studies of body composition by direct chemical analysis of human cadavers are limited. Most of the cadavers were analyzed between 1945 and 1968 and were adults of varying ages who had died because of illness; hence, the values obtained may not be representative of an average healthy adult.

Table 14.1. The contribution of water and protein to the fat-free weights of six adults. From Garrow 1983).

Sex (y)	Water (g/kg)	Protein (g/kg)	Density (kg/m^3	Potassium (mmol/kg)
Male (25)	728	195	1120	71.5
Male (35)	775	165	1083	–
Female (42)	733	192	1103	73.0
Male (46)	674	234	1131	66.5
Male (48)	730	206	1099	–
Male (60)	704	238	1104	66.6
Mean	724	205	1106	69.4
SD	34	28	17	3.3

14.1.1 Applications of cadaver use

Table 14.1 presents data on the contribution of water and protein to the fat-free mass of six adult cadavers. The fat-free tissues of the cadavers were of a relatively constant composition, containing about 72% water and about 20% protein; the potassium content was also relatively constant (about 69mmol/kg). In contrast, the amount of fat was very variable (data not shown), ranging from 4.3% to 27.9% of body weight in the six cadavers.

This page titled 12.2: Chemical analysis of cadavers (14.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.3: Total body potassium (TBK) (14.2)

A constant fraction (0.012%) of potassium exists in the body as the radioactive isotope ^{40}K (half-life = 1.3×10^9y). This isotope emits a high-energy γ-ray of 1.46MeV, allowing the amount of potassium in the body to be estimated by counting with a whole-body γ-spectrometer with sodium iodide detectors (Forbes et al., 1961). The isotope occurs in low concentrations so that the background counts from external radiation (cosmic rays and local sources of ionizing radiation) must be minimized. Hence, the whole-body counter must be shielded from the background radiation with lead, steel, or concrete shielding. Counting times of at least 15min are normally required for adults, with proportionately longer times for children and infants. Such times may be a problem for ill patients. Note that the required equipment is expensive, requires sophisticated technical support: availability is limited.

Calibration of the whole-body counter must be done carefully because the ^{40}K count detected by the whole-body counter is a function of the total body potassium concentration, the geometric configuration of the subject, and internal absorption of the 1.46MeV by the subject. As a result, the counter must be calibrated to allow for differences in the body build of the subjects. Hansen and Allen (1996) achieved this by using a phantom containing a known amount of potassium (as KCl solution) and gender-specific correction factors for weight and height. These authors reported CVs of 1.5% for precision and 4.5% for accuracy for measurements on adults. However, the accuracy of their method has not been confirmed by the analysis of human cadavers.

14.2.1 Application of ^{40}K measurements

Potassium occurs almost exclusively as an intracellular cation, primarily in the muscle and viscera. Negligible amounts occur in extracellular fluid, bone, and other noncellular sites. Measurement of total body potassium can therefore be used as a marker for the body cell mass, and as an index of the fat-free mass in healthy subjects, on the assumption that the fat-free mass has a constant proportion of potassium. Body cell mass represents the total mass of cells in the body that consume oxygen and produce work (i.e., the metabolically active, energy-exchanging mass of the body); it is the nonfat cellular portion of tissues, of which the primary components are skeletal muscle, organ tissue mass, blood, and the brain (Wang et al., 2004). The estimates of fat-free mass generated from total body potassium measurements are accurate and precise at all life stages and in conditions with uncertain hydration status (Naqvi et al., 2018).

Originally ^{40}K measurements were converted from the total body potassium content into the fat-free mass using a value of 69.4mmol K per kg fat-free mass or 2.71g/kg fat-free mass. These values were derived from cadaver analysis. However, a single value for all subjects is now known to be inappropriate and no longer accepted. The potassium concentration of fat-free tissue is a function of age and sex, with both men and women losing on average 5% of their original total body potassium per decade (Figure 14.1). Women have a lower potassium concentration in the fat-free mass than men. Note the reduction in total body potassium with age and the large variation within each age group. The latter emphasizes that total body potassium alone is a poor predictor of fat-free mass unless age- and sex-dependent equations are used (Ribeiro & Kehayias, 2014).

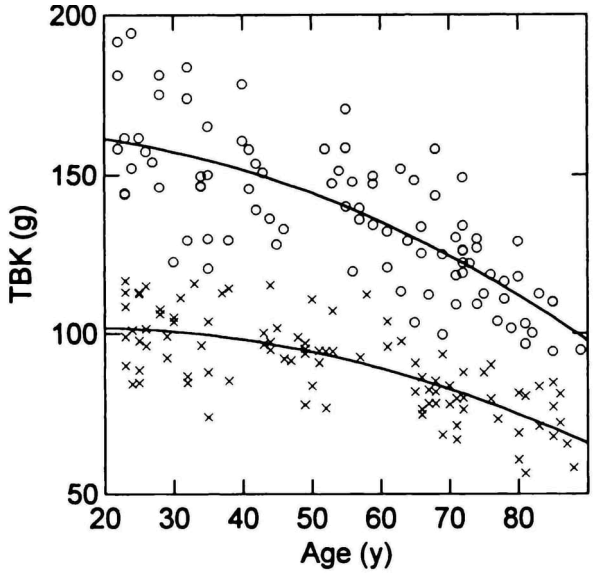


Figure 14.1. Total body potassium (TBK) measurements for females (x) and males (o) as a function of age. From a cross-sectional study with 188 healthy adults. The curves are the best quadratic fit. Modified from Kehayias et al. (1997).

The existence of racial / ethnic differences in total body potassium has also been reported (Ellis et al., 2000). In a recent study, Shypailo and Wong (2020) presented cross-sectional data for total body potassium (in g) measured by whole body counting by sex, race / ethnicity, for six age groups. Black children had higher amounts of total body potassium for certain age groups compared to white and Hispanic children.

The potassium concentration of the fat-free mass also declines with increasing obesity and during pregnancy due to the increased hydration of the fat-free mass. In obese subjects, some of the decrease may also be explained by the lower proportion of muscle in the fat-free mass and by a measurement error resulting from absorption of the γ-rays by adipose tissue (Womersley et al., 1976). During pregnancy, the potassium concentration of the fat-free mass is lower than for non-pregnant women (2.1g/kg vs. 2.3g/kg fat-free mass), but increases in early postpartum (Hopkinson et al., 1997).

Total body potassium may also be altered in clinical patients with a wasting disease such as cancer, when muscle mass is reduced (Cohn et al., 1981). Clearly, constants for the potassium concentration of the fat-free mass must consider age, sex, ethnicity, obesity, and pregnancy, or estimates of fat-free mass derived from ^{40}K measurements will be in error, sometimes by as much as 20%.

The simple 2-component model:

$$\text{Total body fat (kg)} = \text{ body weight (kg) - fat-free mass (kg)}$$

allows the calculation of total body fat (kg) from the fat-free mass (kg). However, any errors and uncertainties in the calculation of the fat-free mass from ^{40}K measurements will be propogated into the derived calculation of total body fat (Silva et al., 2013). These uncertainties can limit the application of the 2-component model in clinical populations with certain disease states, particularly a wasting disease such as cancer, when total body potassium measurements are low as a result of loss of muscle mass. Consequently, in these circumstances, use of the 2-component model to derive indirectly total body fat is inappropriate. Instead, more sophisticated body composition models, involving 3- or 4-components, that reduce the need for underlying assumptions, should be used.

This page titled 12.3: Total body potassium (TBK) (14.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.4: Total body water from isotope dilution (14.3)

Either the stable isotope deuterium (^2H), the radio-active isotope tritium (^3H), or the stable isotope of oxygen (^{18}O) can be used to measure total body water. Of the three, deuterium is now most frequently used. Standardized conditions are necessary for the measurement because fluid and food intake and exercise can all affect total body water concentrations. As a result, samples should be taken in the morning, after an overnight fast and the bladder has been emptied, and with a restriction of fluid intake.

A tracer dose of sterile water labeled with an accurately known amount of the isotope (often deuterium oxide, D_2O) is administered either orally or intravenously to the subject and allowed to equilibrate. Two samples of serum, or urine, or saliva are collected; one prior to the administration of the tracer and a second after the isotope has equilibrated with the subject's water pool (usually 3-5h post-dose). The baseline sample measures the naturally present isotope in the total body water. The enrichment observed in the post-dose sample allows calculation of the total body water. The measurement of enrichment appears to be less reliable in urine than in serum probably because of the longer equilibration period of the bladder contents relative to blood.

Ideally, no food or water is permitted during equilibration, which may take 3-5h, depending on the isotope, the physiological sample, and the health condition of the patient. If fluid has been taken during the equilibrium period, a correction can be made, where necessary, to derive actual body water(Gutiérrez-Marín et al., 2019). Longer equilibration periods are necessary for urine compared to blood serum samples, for obese patients (Schoeller et al., 1980), or for those with edema, ascites, and shock (McMurrey et al., 1958). Table 14.2 demonstrates that the method is relatively robust and, independent of the different physiological fluids used.

Table 14.2. A demonstration of the relatively small variations in total body water (TBW), when calculated from isotopic enrichments of different physiological fluids and at different times postdose.
** Ratio: TBW_A/TBW_B.*
From Schoeller et al. (1985).

Samples (A) & (B)	Isotope	n	*Ratio	SD
(A) Saliva at 4h (B) Serum at 4h	^{18}O	33	1.006	0.019
(A) Urine at 6h (B) Serum at 6h	^{18}O	11	1.012	0.027
(A) Urine at 12h (B) Serum at12 h	^{18}O	14	1.006	0.010
(A) Saliva at 3h (B) Saliva at 4h	^{18}O	20	0.997	0.005
(A) Saliva at 3h (B) Saliva at 4h	^2H	43	0.996	0.007

The calculation of total body water is based on the extent to which the isotopic dose is diluted by the total body fluid as shown below:

$$\text{Total body water } = (V \times C)/(C_2 - C_1)$$

where V = volume of dose, C = concentration of administered isotope, C_1= baseline concentration of isotope in serum / urine / saliva, and C_2 = concentration of isotope in serum / urine / saliva sample after equilibration.

A correction may be necessary for urinary loss of the tracer. Details of the isotope dilution methods and the limitations encountered when measuring total body water in the field have been reviewed by Schoeller (1991) and Ellis (2001), respectively.

The isotopic tracer chosen and the method used for analysis are interrelated. Tritium (^3H) is easy to measure with a scintillation counter but involves radiation to the subject, making the technique unsuitable for children and women of childbearing age (Schoeller et al., 1980), and when repeated measurements over a short time period are necessary. The nonradioactive isotope

^{18}O must be measured by mass spectrometry. The stable isotope deuterium is now the tracer of choice because it is much cheaper than ^{18}O and can be measured by infrared absorption, gas chromatography, an isotope ratio mass spectrometer, and more recently by Fourier transform infrared spectrophotometry (FTIR). Use of FTIR for deuterium (^{2}H) analysis is less expensive and quicker, although slightly less precise than the use of an isotope ratio mass spectrometer. Saliva is preferred to urine as the chosen body fluid when using FTIR for the measurement because urine may cloud the optical lens of the FTIR (Owino et al., 2017). For a comprehensive description of the use of FTIR for the analysis of deuterium in saliva, see:

Recently, an alternative FTIR spectroscopic method (ATR FTIR) has been developed which requires a smaller volume of plasma (10µl), and has shorter analyses times, making it suitable for use in pediatrics; for more details, see Ward (2021).

14.3.1 Application of body water measurements by isotope dilution

The measurement of total body water in both healthy and diseased persons is the most important application of isotope dilution techniques. All body water is present in the fat-free mass. Hence, total body water measurements can be used to estimate the fat-free mass. This estimation requires an assumed value for fat-free mass hydration, as shown below:

$$\text{Fat-free mass (kg)} = (\text{ total body water (kg)})/\text{hFFM}$$

where hFFM = hydration of the fat-free mass.

Based on the 2-component model, once the fat-free mass has been determined, total body fat (TBF) and percentage of body fat can be calculated as shown below:

$$\begin{aligned} \text{Total body fat (kg)} &= \text{ body weight (kg) - fat-free mass (kg)} \\ \text{\% body fat } &= (\text{ Total bodyfat (kg)} \times 100\%)/\text{ body weight (kg)} \end{aligned} \tag{12.4.1}$$

Table 14.3. Median values for hydration by sex. Data from Wells et al. (2010) who also present data for intervening years.

Age	Males Hydration (%)	Females Hydration (%)
5y	76.5	76.7
7y	76.1	75.5
9y	75.7	75.1
11y	75.3	75.0
13y	75.0	74.6
15y	74.4	74.1
17y	73.7	73.7
19y	73.4	73.6

In healthy adults, hFFM is assumed to be 0.732. However, use of a single value for the hydration of fat-free mass may lead to errors in derived values for the fat-free mass. For example, hFFM is known to vary during hormonal cycles in women and in certain disease states owing to under-hydration. Alternatively, over-hydration due to edema, may arise among children with severe acute malnutrition (Gutiérrez-Marín et al., 2019; Most et al., 2018). The hydration of fat-free mass is also known to change during growth and maturation. At birth hFFM is about 80%, after which it gradually decreases until reaching the adult value (Schoeller, 1989). The reference values for hFFM by age and sex in children aged 5-20y compiled by Wells et al. (2010) are shown in Table 14.3, and are similar to earlier values reported for males but about 2% lower for females in mid-childhood (Lohman, 1989).

Hydration values for the fat-free mass are also affected by obesity, the increase said to be due to an expansion of extracellular water, although other mechanisms may be involved (Leone et al., 2000); see Chumlea et al. (2007) for more details. Note the marked increase in hFFM (as %) in heavier BMI groups shown in Figure 14.2 based on a large study of subjects from 4-22y (Gutiérrez-Marín et al., 2019). This same trend has been reported previously in smaller studies of obese children (Haroun et al., 2005; Wells & Fewtrell, 2006) and adults (Waki et al., 1991), and highlights that reference data for hydration values for higher BMI groups are also needed to avoid bias.

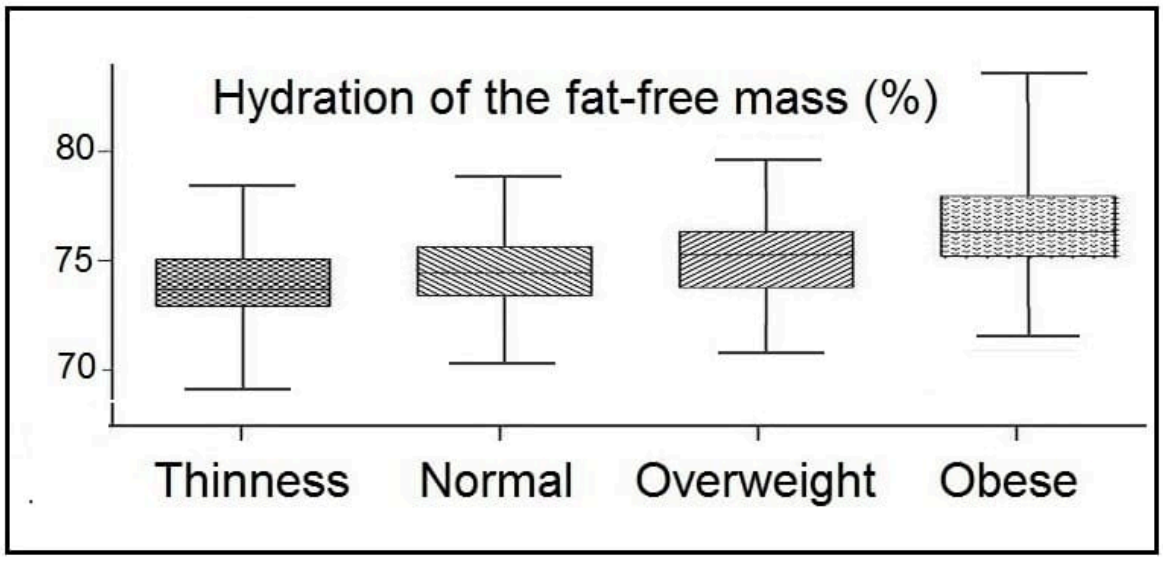


Figure 14.2. Values for hydration and density of fat-free mass based on the four-compartmental model and stratified by nutritional status grouped by BMI SD score for UK subjects aged 4-22y. Redrawn and abbreviated from Gutiérrez-Marin et al. (2019).

Pregnancy is also associated with an increase in hFFM, as noted in Section 14.2. Published values for hFFM at specific time points in pregnancy are shown in Figure 14.3.

The equations of van Raaij et al. (1988), based on ten-week intervals throughout gestation, are those most used in 2-component models to estimate maternal fat mass during pregnancy. A regression equation based on the van Raaij et al. (1988) data has been developed allowing estimation of hFMM for any given time throughout gestation; this regression line is also shown in Figure 14.3.

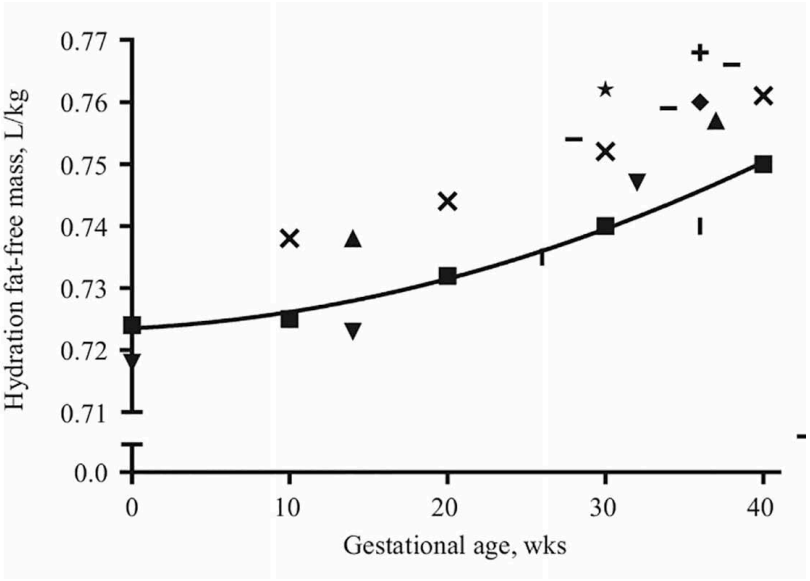


Figure 14.3. Published values for FFM hydration throughout pregnancy in published studies. The exponential regression line is based only on the data by van Raaij et al.(1988). Redrawn from Most et al. (2018) who also provide information on the sources of the individual data points.

Table 14.4 highlights the effect of adopting different hydration constants on the assessment of fat-free mass and fat mass in a women at 30wks gestation with a normal pre-gravid BMI (i.e., 24.8kg/m^2). Clearly, failure to consider the increase in hFFM in models based on total body water using isotope dilution would result in an overestimate of the actual fat-free mass in pregnant women, and thus underestimate fat mass when the 2-component model is applied.

Table 14.4. Effect of adopting different hydration constants on fat mass (FM) and fat-free mass (FFM) calculated at 30 weeks gestati (45.0L). Data from Most et al. (2018).

Author	Hydration constant	FFM (kg)	FM (k
Siri (1961)	0.724	62.2	13.8
Van Raaj et al. (1988)	0.740	60.8	15.2
Fidanza (1987).	0.752	59.8	16.2
Catalano et al. (1995)	0.762	59.1	16.9

Recently, reference charts for infants and children aged from 6wks to 5y by sex have been developed for percentiles for total body water measured by deuterium dilution, and fat-free mass calculated from total body water using published hydration coefficients (Wells et al., 2020). These charts will help clinicians identify how specific diseases and their treatment impact on fat-free mass and fat mass in this age group.

The error associated with the measurement of total body water using these isotopic tracers is typically less than 1kg. This error equates to an uncertainty of about 10% (about 1.4kg) in the absolute fat mass of an average healthy adult or 2% in the estimate of percentage of fat. These relatively low errors have resulted in the isotope dilution method becoming the reference or gold-standard method in comparison with other measurement procedures for total body water (e.g., bioelectrical impedance: Section 14.8).

A 3-component model that partitions body weight into three major components can also be used to measure fat-free mass and fat mass:

$$\text{Body mass} = \text{fat} + \text{water} + \text{residual}$$

where residual (i.e., fat-free dry mass) is the sum of protein, bone mineral, and glycogen

This 3-component model partitions the body mass into fat, total body water, and the remaining fat-free dry mass, which is assumed to have a constant ratio of protein to mineral. This 3-component model avoids the assumption that the water content of fat-free mass is the same for all individuals of a given age and sex. Instead, the model provides an estimate of the hydration and density of fat-free mass. Three measurements are made: body weight (in kg), body volume (in liters), and total body water (in liters). Fat mass can be calculated from these three basic measurements by applying the 3-component model of Siri (1961) which includes both whole body density and total body water; see Wells et al. (1999) and Fuller et al. (1992) for further details. This version of the 3-component model provides improved practical estimates of both the fat mass and fat-free mass in children (Silva et al., 2013).

This page titled 12.4: Total body water from isotope dilution (14.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.5: Multiple dilution methods (14.4)

Multiple dilution can be used to estimate the volume of various body fluid compartments that, in turn, can be used to estimate two components of the fat-free mass: extracellular mass (ECM) and the body-cell mass (BCM).

Typically, multiple dilution involves determining both total body water (via isotope dilution; Section 14.3) and extracellular water (ECW), the latter using a tracer such as bromide that does not enter the intracellular space (Wong et al., 1989). The difference between these two measurements (i.e., TBW – ECW) reflects the intracellular water (ICW), which is more metabolically active than ECW and provides an estimate of BCM (Ribeiro & Kehayias, 2014).

To estimate BCM using this approach, a pre-dose blood or urine sample is taken followed by dosing with deuterium oxide and sodium bromide solution. After a 3-5h equilibration period, a final post-dose blood or urine sample is collected, although longer equilibration periods are needed for individuals with expanded ECW, including those with extreme obesity. Deuterium enrichment of the biological sample is described in Section 14.3, whereas bromide enrichment is determined by high-performance liquid chromatography or non-destructive liquid X-ray fluorescence. The reported precision for ICW calculated using this approach is about 2.5% provided standardized protocols are used (Earthman, 2015). The ratio of ECW / TBW can also be calculated using this approach. In a study of nursing home elderly, the ratio ECW / TBW was significantly higher compared to the ratio in free-living elderly, prompting the suggestion that this ratio may have potential as a surrogate method for the clinical assessment of frailty (Kehayias et al., 2012).

14.4.1 Application of multiple dilution methods

The ECM is defined as the component of the fat-free mass which exists outside the cells. It consists of both fluid (e.g., extracellular fluids, plasma volume) and solid (e.g., skeleton, cartilage, tendons) components which are involved in transport and support and are not metabolically active.

In contrast, the BCM is the total mass of cells in the body that consume oxygen and produce work (i.e., the metabolically active, energy-exchanging mass of the body). These components are the nonfat cellular portion of tissues, primary the skeletal muscle, organ tissue mass, blood, and the brain (Wang et al., 2004). Nutritional status, physical activity level, and disease states alter BCM, which thus serves as a helpful biomarker.

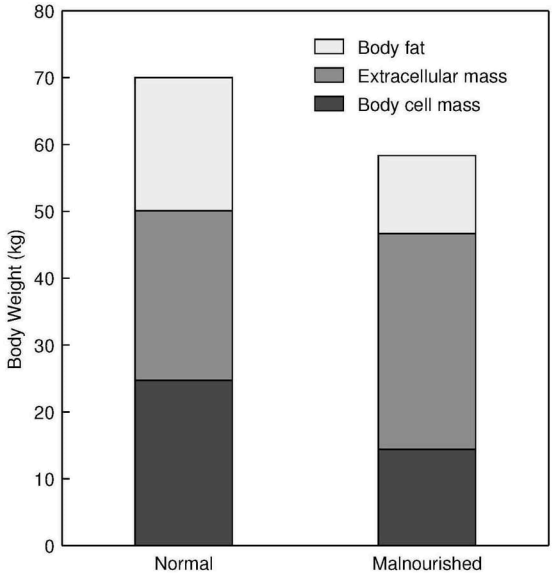


Figure 14.4. The mean body composition of 25 normally nourished healthy volunteers and 75 malnourished patients. From Shizgal (1981).

Measurements of BCM and ECM are especially critical in malnourished patients. Values for total fat-free mass in these patients may remain unchanged, but the composition of the fat-free mass is abnormal, with a reduced BCM, concomitant with an expansion of the ECM, as shown in Figure 14.4. Hence, any loss in body weight in such patients reflects a loss of body fat (Shizgal, 1987).

This page titled 12.5: Multiple dilution methods (14.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.6: In vivo activation analysis (14.5)

A group of related techniques involving in vivo neutron activation analysis (NAA) allow the direct estimation of the amount of a range of chemical elements in the living human body. Most other techniques used in body composition studies, with the exception of whole body counting for potassium, generate data on tissue density or volume, but not data on the amount of a component. As a result, multicomponent elemental models based on in vivo NAA have gradually become accepted as reference methods for the calibration of many of the other techniques described in this chapter.

Nearly all the major elements present in the body can be analyzed by in vivo NAA, including hydrogen, oxygen, carbon, nitrogen, calcium, phosphorus, sodium, and chlorine (Cohn et al., 1984). Of special interest is the application of NAA to measure the carbon-to-oxygen (C/O) ratio in vivo . This ratio can provide an independent, unbiased measure of the distribution of fat and fat-free mass, which is not dependent on the assumptions about the composition of fat-free mass. Small changes in fat-free mass can be monitored, making the method appropriate for studying the depletion of fat-free mass with aging (Kehayias et al., 2000).

A major negative factor associated with in vivo NAA is that the subject is exposed to radiation. This and the associated risks must always be explained to the subject. In vivo NAA is most used in clinical medicine for the determination of total body nitrogen (TBN), total body calcium (TBCa), and bone mass as discussed below. TBN and TBCa are only determined in a relatively small number of laboratories worldwide — an indication of the expense involved and technical difficulties associated with the method.

14.5.1 Total body nitrogen by in vivo NAA

Nitrogen is normally determined by NAA by bombarding the patient, in a supine position, with a low neutron flux from a ^{238}PuBe source or from a cyclotron or neutron generator. During irradiation, a proportion of ^{14}N is converted to an excited state of ^{15}N, which decays almost immediately to its ground state, emitting a “prompt” γ-ray at 10.83MeV. This activity is counted by an array of sodium iodide detectors in a whole body counter (Figure 14.5).

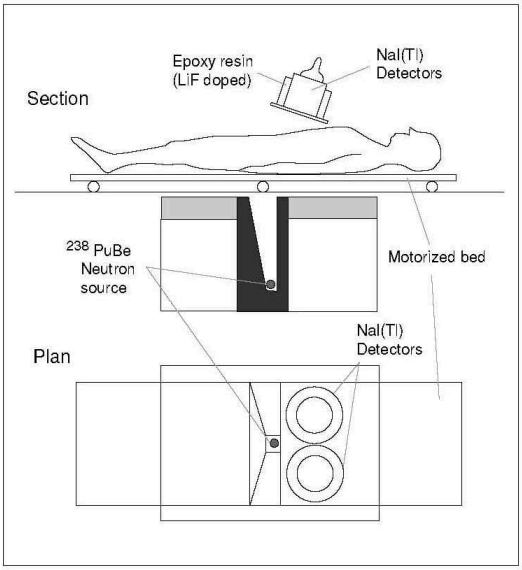


Figure 14.5. in vivo γ-neutron activation. Redrawn from Cohn et al. (1981b).

The detected γ-ray counts are proportional to the absolute mass of total body nitrogen (Beddoe & Hill, 1985). Measurement of the nitrogen content of the body gives a measure of total body protein because the mass of nitrogen bears a fixed ratio to the mass of protein (1g N = 6.25g protein). The radiation exposure using this method is about 0.3mSv. This compares with the national background radiation from cosmic and other sources of about 3.5mSv/y.

The 10.83MeV γ-ray is specific to nitrogen and is at an energy which is not affected by interference from other reactions. Extensive shielding, however, is necessary around the sodium iodide detectors to reduce the level of background radiation. Nevertheless, corrections to the γ-count must still be made for the background.

The calibration of the counter is critical and must consider the height, weight, and adiposity of the subject because of the varying neutron attenuation and internal absorption of the emitted γ-rays. Calibration is achieved using phantoms or by using hydrogen as an internal standard Vartsky et al., 1979).

The accuracy of prompt γ-neutron activation for measuring total body nitrogen has been validated by comparing total body nitrogen in two human cadavers with results obtained by direct chemical analysis of nitrogen (Knight et al., 1986). Close agreement between the two techniques was found. This study also confirmed the use of the ratio 6.25 for the relationship between total body protein and total body nitrogen.

14.5.2 Application of body nitrogen by in vivo NAA

Changes in total body protein of hospital patients with diseases such as cancer, renal dysfunction, hypertension, chronic heart disease, and rheumatoid arthritis, and with severe trauma or sepsis have been studied using prompt γ-neutron activation (Beddoe & Hill, 1985). The results indicate that substantial losses of body protein may occur, even when conventionally adequate nutritional support has been provided for some of these patients. Depleted total body protein is also a consistent finding among acutely ill anorexic patients. In a longitudinal study, Haas et al. (2018) reported that after recovering from anorexia nervosa, depletion of body protein measured by in vivo INAA remained in adolescent patients after 7mos, even though body weight was restored. Clearly, further work is required to identify nutritional intervention procedures which minimize loss of body protein in hospital patients and those with anorexia nervosa.

14.5.3 Total body calcium by in vivo NAA

The method is based on the conversion of a proportion of the naturally occurring isotope ^{48}Ca in the body to ^{49}Ca (half-life = 8.8min) by exposing the patient to a low neutron flux. Immediately following irradiation, the patient is transferred to a whole-body counter, and the γ-rays emitted by the decay of ^{49}Ca are detected by an array of sodium iodide detectors.

The counting geometry is similar to that used for total body potassium, and again care must be taken in positioning the subject and correctly accounting for the varying height, weight, sex, and body mass index of subjects. In particular, for subjects with BMI > 30, the effects of neutron attenuation become significant, necessitating additional corrections or a special calibration (Ma et al., 2000). The reported accuracy and precision are from 1-2% (Cohn et al., 1974).

14.5.4 Application of total body calcium by in vivo NAA

The radiation exposure using this method varies from 2.5-25mSv, depending on the neutron source. This range is considerably higher than that experienced during dual X-ray absorptiometry (DXA) and limits the general applicability of the method. Consequently, the use of neutron activation analyses for the assessment of total body calcium has largely been replaced by DXA (Section 14.11). Instead, the method is now mainly used as a calibration tool for other techniques.

The assessment of total body calcium by in vivo neutron activation analysis is also discussed in Chapter 23 under the assessment of calcium status.

This page titled 12.6: In vivo activation analysis (14.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.7: Densitometry (14.6)

Body density was one of the first measures of body composition to be made Behnke et al., 1942). It is relatively easy to measure, and in the past, body density was the gold-standard method to determine percentage body fat using the 2-component model. Certain disease states characterized by excess fluid retention and under-mineralization decrease the density of fat-free mass. Consequently, densitometry is now often combined with other measures in a 4-component model of body composition so that hydration and density of fat-free mass are measured along with the actual bone mineral content, thus providing a more accurate assessment. (Silva er al., 2013).

Hydrostatic weighing was the initial densitometric method used to determine body volume, and hence whole body density. However, it is now being replaced by plethysmographic methods that are more acceptable to subjects, particularly children (Fields & Goran, 2000; Wells & Fewtrell, 2006). All three methods are described below. A concluding section describes the calculation of body fat from body density (Section 14.6.4).

14.6.1 Hydrostatic weighing

The conventional method of directly measuring whole body density involves weighing the subject and then using Archimedes' principle to determine the volume of the subject. Thus, the subject is weighed first in air and then when completely submerged in water in a large tank. The subject is instructed to squeeze out any air bubbles trapped inside the bathing suit, and to expel as much air as possible from the lungs before immersion. The hydrostatic weight is recorded at the end of the forced expiration. Multiple readings should be taken using a continuous and sensitive recording of underwater mass, the heaviest corresponding to the most complete expiration. This method requires a high degree of water confidence and thus the method is not suitable for children younger than 8y, the elderly, obese, or unhealthy persons.

The body volume is then calculated from the apparent loss of weight in water (i.e., the difference between the weight of the person in air and his or her corresponding weight in water). Once total body mass and body volume have been determined, whole body density can be readily calculated, on the basis that density is mass per unit volume and the density of water is l.0kg/L at 4°C:

$$\text{Whole body density } = (\text{ body weight in air (kg)})/(\text{ apparent loss in weight (kg)})$$

Three corrections must be applied:

- Hydrostatic weighing is usually performed in water at 30°C instead of 4°C. At this higher temperature, the density of water is 0.9957kg/L, and a water temperature correction factor must be applied.
- Air trapped in the lungs also contributes to the amount of water displaced by the subject under water. Residual air in the lungs can be measured while the subject is in the tank. Alternatively, the volume can be estimated using a nitrogen washout, helium dilution, or oxygen dilution (Brodie, 1988). This correction can also be estimated from empirical equations. The residual air volume is then subtracted from the body volume.
- The volume of the air trapped in the gastrointestinal tract also contributes to the total amount of water displaced. This volume is small and is never measured. It is often taken as 100mL.

The within-subject variability in gastrointestinal gas volume, however, can be quite large (0-500mL in adults), reducing the precision of the method. Even when the above corrections are carefully applied, considerable uncertainty remains. If the whole body density is used to calculate the percentage of body fat (see below), these uncertainties become errors in the percentage of body fat. The errors may be systematic rather than random in nature and may be large when the residual air volume is estimated. If the residual air volume is in error by 300-500mL, there will be a corresponding uncertainty in the percentage of body fat from 3-5%.

Hydrostatic weighing can give very reproducible results for whole body density, provided that the examiners and the subjects are well trained. For example, Durnin and Rahaman (1967) reported a standard deviation of 0.008kg/L for serial measurements on three subjects over a 1y period.

14.6.2 Water-displacement plethysmography

The use of a plethysmograph eliminates the necessity for totally immersing the subject in water, a disadvantage of hydrostatic weighing (Section 14.6.1). For the measurement, the plethysmograph is zeroed and filled with water (Figure 14.6).

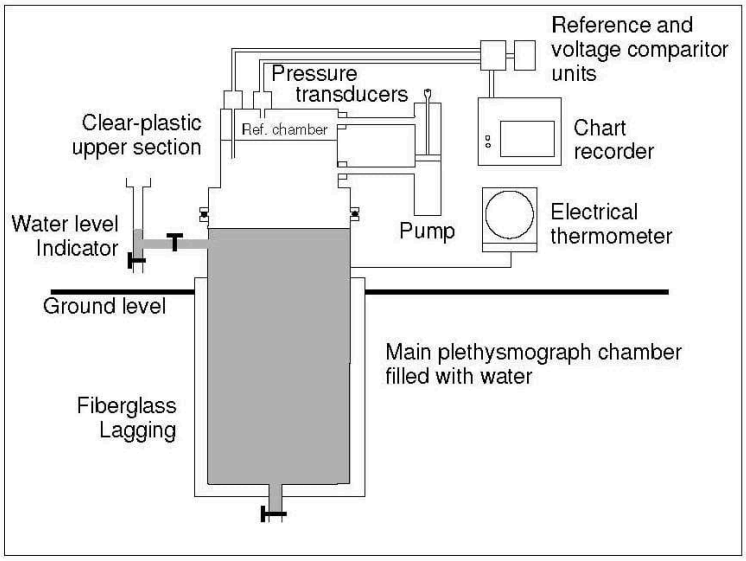


Figure 14.6. Measurement of body density using a plethysmograph. From Garrow et al. (1979)

The subject is then weighed, and a weight of water equal to the weight of the subject is removed from the plethysmograph. The subject then stands with water up to the neck only, and the head is covered by a clear-plastic dome. The volume of air surrounding the head of the subject, and in the lungs and gut, is then determined by measuring the pressure changes produced by a pump of known stroke volume (Garrow et al., 1979). This allows the total volume of the subject to be determined. The total time for the test, including three measurements on each subject, is about 20min.

The method has been used successfully to measure body density of obese adults (Garrow et al., 1979). Estimates of body fatness obtained compared favorably with those based on total body potassium.

Water-displacement plethysmography has now been replaced by air-displacement plethysmography (discussed below) for the measurement of total body volume. Nevertheless, water-displacement plethysmography has been adapted to measure leg volume, and is used by clinicians to study chronic venous insufficiency (CVI). For a review of the clinical use of water displacement leg volumetry and the potential errors that may occur, see Rabe et al. (2010).

14.6.3 Air-displacement plethysmography

The measurement of body volume has been significantly eased by the development of an air-displacement plethysmography device (Bod Pod). These devices determine the volume of a subject indirectly by measuring the volume of air displaced by the subject inside an enclosed chamber — plethysmograph or Bod Pod. The method is quick, comfortable, automated, non-invasive, safe, and does not require extensive technical training. For infants up to about 6mos of age, a device known as the “Pea Pod” is available and can accommodate infants up to about 8kg. For small children from 2-6y, the Bod Pod with a Pediatric Option can be used, whereas the standard Bod Pod can be used for older children (i.e., > 6y), adults, and the elderly.

The air-displacement plethysmograph consists of an ovoid fiberglass structure divided into two sections: a rear reference chamber and a front test chamber containing the seated subject (Figure 14.7).

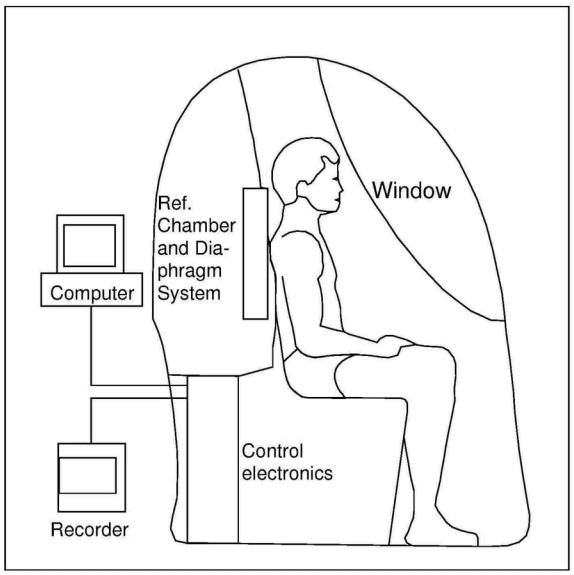


Figure 14.7. Measurement of body volume using an air-displacement plethysmograph. Modified from Dempster & Aitkens (1995).

The dividing wall between the two chambers encompasses a large diaphragm. This can be made to oscillate under computer control, generating complementary pressure changes in the test and reference chambers which are recorded. These changes and the application of the basic gas laws allows the volume of the test chamber to be calculated. Measurements both with and without a seated subject provide information on the total body volume of the subject.

While they are seated in the test chamber, subjects should wear minimal clothing, often a swimsuit and a tightly fitting bathing cap to minimize the volume of air contained near the hair or skin, and in clothing. Measurements over a 1-min period usually suffice, and an average of two separate trials should be used to calculate body volume. For details on the methods and performance of the PEA POD system for measuring body composition in infants, see Ellis et al. (2007); for moderately premature infants, see Forsum et al. (2016), and for children aged 2-6y, see Fields et al. (2012).

A correction for the average volume of air in the lungs and thorax during normal breathing (VTG) should be applied. This can be measured while the subject is in the test chamber (Dempster & Aitkens, 1995), or predicted based on age, sex, and height (McCrory et al., 1998). Pregnancy must also be considered because thoracic gas volume declines throughout pregnancy by about 100mL per trimester. Hence, VTG needs to be adjusted for the trimester-specific decline in lung volume (Most et al., 2018). Once body volume has been measured, whole body density (D) can be calculated from measured body mass.

For a comparison of the components and capabilities of the three plethysmography techniques, see Fields et al. (2015).

14.6.4 Application of densitometry to calculate body fat

Once whole-body density has been measured by one of the methods outlined above, the percentage body fat can be calculated. This involves the selection of an empirical densitometric equation relating fat content to whole body density (D). Several empirical densitometric equations have been derived based on the classic two-component model for body composition. As noted earlier, in this model body weight is divided into fat and fat-free mass, and relies on assumptions that ignore inter-individual variability in the composition of the fat-free mass.

All the classical densitometric equations shown below assume:

- The density of the fat-free mass is relatively constant.
- The density of fat for normal persons does not vary among individuals.
- The water content of the fat-free mass is constant.
- The proportion of bone mineral (i.e., skeleton) to muscle in the fat-free body is constant.

The different authors all assume the density of fat and the fat-free mass do not change with age and sex. There are two constants in each equation, $C_1$and C_2, the different authors suggesting slightly different values for each:

$$\begin{aligned} \%\ \mathrm{F} &= ((4.950/\mathrm{D}) - 4.500) \times 100\%(\ \mathrm{Eq.A}\ ) \\ \%\ \mathrm{F} &= ((4.570^*/\mathrm{D}) - 4.142) \times 100\%(\ \mathrm{Eq.B}) \\ \%\ \mathrm{F} &= ((5.548^*/\mathrm{D}) - 5.044) \times 100\%(\ \mathrm{Eq.C}) \end{aligned} \tag{12.7.1}$$

The Siri (1961) equation (Eq.A) is most widely used and assumes that the densities of fat and fat-free mass are about 0.90 and 1.10kg/L respectively. Brožek et al. (1963) (Eq.B) and Rathbun and Pace 1945) (Eq.C) used the concept of a reference man of a specified density and composition; these equations avoid the requirement of estimat;ing the density of fat-free mass. The constants applied in all these equations came from the chemical analysis of a few adult cadavers dissections, animal data, and indirect estimates of fat-free mass in human subjects (Heymsfield et al., 1991; Silva et al., 2013).

None of these classical empirical equations relating fat content to body density, however, are suitable for adult patients in clinical settings when the composition of their fat-free mass may be abnormal. This will include patients undergoing hyperalimentation with high-sodium fluids, or with congestive heart failure or liver disease, as total body water content as a fraction of fat-free mass may be markedly higher in these patients, thus violating the assumption that the water content of the fat-free mass is a constant (i.e., 73.2%). In these patients, the density of fat-free mass is decreased. Not surprisingly, in patients with diseases associated with under-mineralization, the density of fat-free mass is also decreased (Werdein & Kyle, 1960). Consequently, in all these patients, fatness will be overestimated when the 2-component model is applied (Wells & Fewtrell, 2006). More recent research has raised concerns over the assumption of constant properties for hydration and density of fat-free mass when these classical empirical equations are applied to assess body composition not only in patients with certain diseases, but also in healthy children and adolescents, the elderly, pregnant women and those with obesity, as noted earlier.

Although fat has relatively uniform properties throughout the life course (zero water and a density of 0.9007kg/L), in contrast fat-free mass has different properties in children compared to adults. This arises because of chemical maturation of the fat-free mass during growth which results in higher levels of water and lower levels of mineral and proteins, and thus changes in the hydration and density of the fat-free mass (Wells et al., 2010), as noted earlier. Nevertheless, the adult-derived values for the density and hydration of fat-free mass and applied in the classical equations shown above have often been used to study body composition in children (Silva et al., 2013).

In an effort to improve the accuracy in the estimates of percentage body fat based on pediatric densitometry (and hydrometry, Section 14.3, Wells and co-workers measured body composition via a 4-component model in a large, healthy sample of children and adolescents aged 4-23y. Use of this model overcomes the limitations associated with the assumptions of constant properties for hydration and density of fat-free mass applied in the Siri equation.

Table 14.5. Median values for hydration, density, and constants (C_1 and C_2) for the pediatric version of the equation of Siri (1961), obt (2010).

	Males				Females
Age	Hydration (%)	Density (kg/L)	C_1	C_2	Hydration (%)
5y	76.5	1.0827	5.36	4.95	76.7
6y	76.3	1.0844	5.32	4.90	76.1
7y	76.1	1.0861	5.28	4.86	75.5
8y	75.9	1.0877	5.24	4.82	75.2
9y	75.7	1.0889	5.21	4.79	75.1
10y	75.5	1.0900	5.19	4.76	75.0
11y	75.3	1.0911	5.16	4.73	75.0
12y	75.2	1.0917	5.15	4.72	74.9
13y	75.0	1.0920	5.14	4.71	74.6
14y	74.8	1.0927	5.13	4.69	74.4

	Males				Females
Age	Hydration (%)	Density (kg/L)	C_1	C_2	Hydration (%)
15y	74.4	1.0942	5.09	4.66	74.1
16y	74.0	1.0960	5.05	4.61	73.8
17y	73.7	1.0978	5.02	4.57	73.7
18y	73.5	1.0991	4.99	4.54	73.6
19y	73.4	1.1000	4.97	4.52	73.6
20y	73.3	1.1006	4.96	4.51	73.6

Their results confirmed that the previously published adult values for the density and hydration of fat-free mass when applied to pediatric populations were inappropriate. They developed new empirical age- and sex-specific reference values for the hydration and density of fat-free mass; these are shown in Table 14.5. In addition, they developed prediction equations of density and hydration for fat-free mass by age, sex, and body mass index (BMI) SD score; see Wells et al. (2010) for more details.

Note that the values for the C_1 and C_2constants for the adult males (age 20y), shown in Table 14.5, are similar to the corresponding values shown in the Siri equation above (i.e., 4.95 for C_1 and 4.50 for C_2), whereas for adult females, the corresponding values are slightly lower: 4.90 for C_1 and 4.44 for C_2 at 20y. With the substitution of the age- and sex-specific $C_1$and C_2 constants from Table 14.5 in place of the C_1 (4.95) and C_2 (4.50) constants in the Siri equation, the accuracy of the 2-component model for estimating fat mass of a healthy pediatric population based on densitometry could be improved.

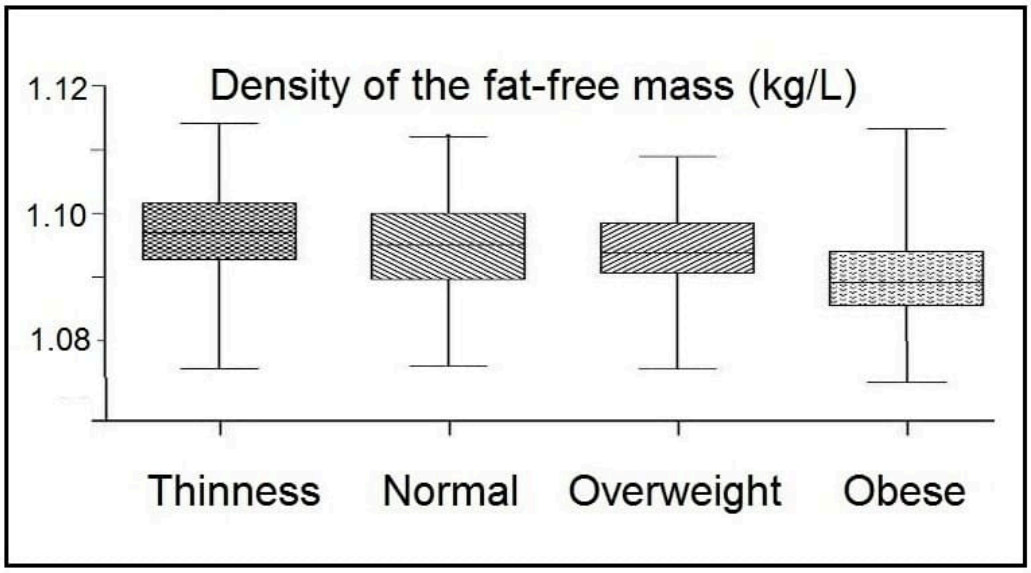


Figure 14.8. Distribution of values for density of fat-free mass based on the 4-component model and stratified by nutritional status grouped by BMI SD score for UK subjects 4-22y. Redrawn and abbreviated from Gutiérrez-Marin et al. (2019).

More recent research has highlighted increasing values for hydration of fat-free mass (see Section 14.3 and Figure 14.3) but decreasing values for the density of fat-free mass in the children with a heavier body mass index (BMI) (Figure 14.8). This trend was observed earlier in children (Haroun et al., 2005) and adults (Waki et al., 1991). Gutiérrez-Marin et al. (2021) have developed a predictive equation to estimate the density of fat-free mass for use when using a 2-component model to assess children and adolescents with different degrees of obesity. Application of the equation shown below improves the accuracy and precision of the density of the fat-free mass (DFFM) estimates in this population.

$$\mathrm{DFFM} = 1.0791 + (0.009 \times \text{ age } + 0.0021 \times \text{ gender }) - (0.0014 \times \text{ BMISDS })$$

Where age is in years; gender 1 (male), 2 (female); DFFM = density fat-free mass; BMISDS = body mass index Z-score

An excel file with all the steps in the calculation is available online as supplementary material in Gutiérrez-Marin et al. (2021).

Similar trends in the values for the hydration and density of fat-free mass have been observed in studies of older Hispanic Americans (González-Arellanes et al., 2019) and obese Mexicans (González-Arellanes et al., 2021) all aged > 60y. Again values for fat-free mass density were lower, but higher for the hydration fraction suggesting that modifying the assumptions regarding both the density and hydration values for fat-free mass applied in the classical densitometric empirical equations may also be appropriate for the elderly and in conditions of obesity. Once percentage body fat has been calculated, then total body fat (TBF) and/or fat-free mass can be derived as follows:

$$\begin{aligned} \text{Total body fat (kg)} &= \text{ body wt (kg)} \times \% \text{ body fat } \div 100 \\ \text{Fat-free mass (kg)} &= \text{ body wt (kg)} - \text{ total body fat (kg)} \end{aligned} \tag{12.7.2}$$

This page titled 12.7: Densitometry (14.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.8: Total body electrical conductivity (14.7)

Total body electrical conductivity (TOBEC) is measured by observing the changes induced by placing the subject in an electromagnetic field Baumgartner, 1996). The extent of the change depends on the overall electrical conductivity of the body and, in particular, on the proportions of fat and the fat-free mass in the body: the fat-free mass, comprised largely of water with dissolved electrolytes, will readily conduct an applied electric current, whereas fat is anhydrous and a poor conductor. In more modern equipment, the subject lies supine on a motorized bed (Figure 14.9) that is passed in a series of steps progressively through a uniform solenoid coil.

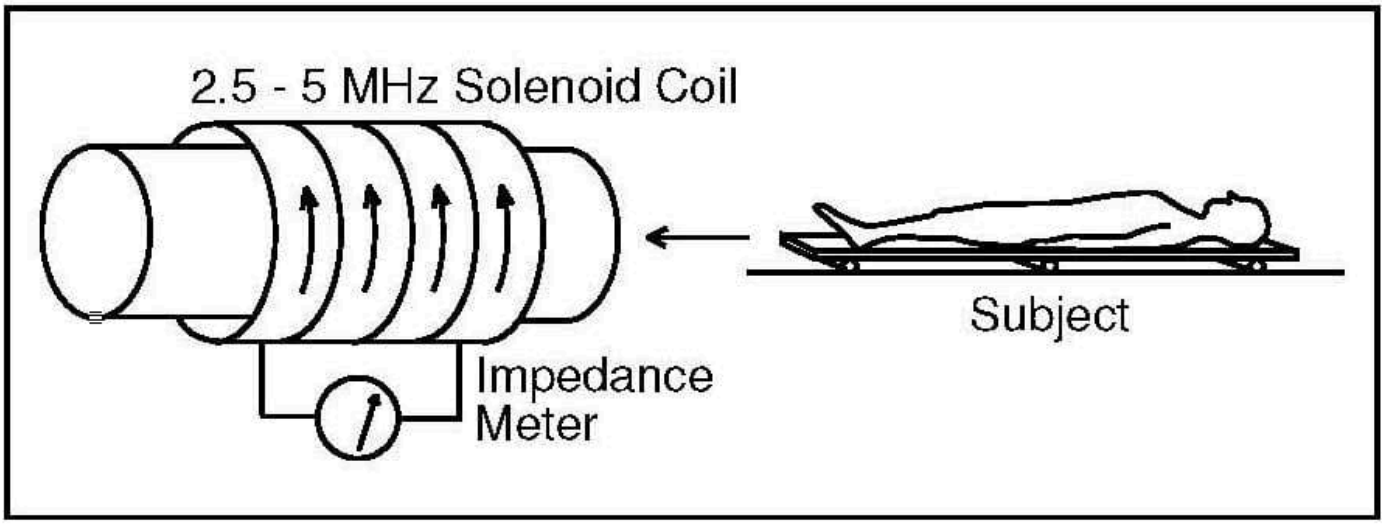


Figure 14.9. Measurement of body composition by total body electrical conductivity. Modified from van Itallie et al. 1985).

A 2.5MHz oscillating radiofrequency current is passed through the coil. This induces an electromagnetic field in the space enclosed by the coil, which, in turn, induces a current in the subject, the magnitude depending on the conductivity of the subject. A second reading is taken when the coil is empty; the difference represents the measurement. The use of Fourier analysis improves the assessment of the lean body mass (Van Loan & Koehler, 1990).

Measurement of TOBEC is simple, safe, noninvasive, fast, and causes no discomfort to the subjects. The technique can be used on individuals who cannot be weighed underwater. However, the instrument is expensive.

Careful consideration must be given to the positioning of the subject and corrections for body geometry and length applied. Treuth et al. (2001)used a variety of techniques, including TOBEC, to determine body composition in 8y prepubertal girls. The authors concluded that the determinations are highly method dependent and that results from all six methods studied (TOBEC, total body potassium, isotope dilution for total body water, bioelectrical impedance, anthropometry, and DXA) are not interchangeable. The results highlight how TOBEC measurements can be evaluated by comparison with other methods. Unfortunately, most of the other methods are also subject to error and bias.

14.7.1 Applications of total body electrical conductivity

To interpret TOBEC measurements in terms of the quantity of fat-free mass in the body, a calibration equation is applied generated by measuring the fat-free mass of a reference population using an alternative technique and relating this to the TOBEC value of each individual. A calibration equation has also been developed to measure fat-free mass and fat reliably and precisely in neonates through 1y of age (Fiorotto & Klish, 1991; Fiorotto et al., 1995).

TOBEC is relatively insensitive to shifts of fluid or electrolytes between the intracellular and the extracellular compartments and to variations in bone mineralization. For example, in a study of middle aged and elderly subjects (35-90y), higher values for fat-

free mass were reported when predicted from measurements based on TOBEC compared to those for fat-free mass predicted by using densitometry or hydrometry (Van Loan & Koehler, 1990). These findings led to the suggestion that the higher fat-free mass values arise because the TOBEC signal is unaffected by decreases in bone mineralization.

This page titled 12.8: Total body electrical conductivity (14.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.9: Bioelectrical impedance (14.8)

The use of bioelectrical impedance (BIA) for the assessment of body composition is widespread, in large part due to the affordability, portability, and ease of use of the bioimpedance devices. None of the BIA devices measure body composition directly. They measure an electrical response of the body, resistance, when exposed to an electrical current. The resistance measured is transformed into a prediction of total body water by an algorithm, and from that the body fat mass is determined.

Three categories of BIA devices are available commercially: single-frequency (SF-BIA), multiple-frequency (MF-BIA), and bioimpedance spectroscopy (BIS); these are described below together with some new developments in BIA. Both the SF-BIA and MF-BIA devices rely upon population-specific prediction equations. In contrast, BIS uses biophysical modeling to estimate body compartments. All the algorithms are based on or validated against other body composition reference methods, which are not totally accurate and error-free. Each BIA device works with an inbuilt algorithm specific to the device and for the population for which the algorithm was developed so it is not possible to compare studies unless the same combination of device / equation / population is used (Ward, 2019; Sheean et al., 2020). Furthermore, the reference population on which the BIA algorithm was based must be appropriate for the target subject being measured (Lemos & Gallagher, 2017).

Many factors may influence the precision and accuracy of BIA techniques. They include factors associated both with the individual (e.g., degree of adiposity, fluid and electrolyte status, skin temperature) and with the environment (ambient temperature, proximity to metal surfaces and electronic devices), the assumptions underlying prediction or modeling equations, instrumentation, and variations in the protocols used for the BIA measurements. Box 14.2 presents recommendations for optimizing whole body BIA measurements.

Box 14.2. Recommendations for optimizing whole body BIA measurements in a adult. From Earthman (2015).

Preparing for the measurement

- **Food/beverage and activity.** Individual should fast (no food that day except water) and avoid alcohol, caffeine, and exercise for at least 8h prior to measurement in the morning (research settings); shorter time frames and other times of day may be acceptable in the clinical setting — note time of day for consistency in follow-up measures.
- **Void bladder**. Individual should void bladder prior to measurement.
- **Clean skin surface**. Clean skin surface well with alcohol; individual should not use lotion or oils on the skin prior to measurement; avoid placing electrodes on broken skin.
- **Device calibration**. Calibrate the bioimpedance device according to the manufacturer's recommendations prior to measurement.
- **Height and weight**. Obtain an accurate measure of height and weight.

Testing conditions and considerations

- **Device placement**. Place device on a nonmetal surface, at least 1m away from electronic or magnetic devices.
- **Ambient temperature**. Avoid excessively warm or cool ambient temperature.
- **Electrodes and leads**. Use electrodes with sufficient surface area (≥ 4cm^2); store electrodes in sealed bag away from heat; use device-specific leads provided by manufacturer.
- **Electrode placement**. Place electrodes at least 5cm apart, if possible; proximal electrodes should never be moved from standard anatomical site placement; if necessary, the distal electrodes may be moved to achieve at least 3cm of separation; the most important thing is to measure and record distance between electrodes to ensure placement consistency for follow-up measurements.
- **Side of body**. If using standard tetrapolar placement of electrodes, measure on the same side of the body as previous measures; in individuals with amputations, muscle atrophy, or other abnormal conditions, use the nonaffected side, if possible; be consistent on side of measurement for follow-up. Right-side measurements are commonly used in the literature.

- **Body positioning and limb separation**. Body position should be supine, except for stand-on scale devices, with arms separated ≥30° from the trunk and legs separated by about 45°; in individuals with overweight and obesity, separate arms from trunk and legs from each other using rolled cotton towels/blankets.
- **Fluid and electrolyte status**. Note if serum electrolytes are abnormal; it is best to conduct bioimpedance measurements only when serum electrolytes are normal. Note if edema is present; edema causes lower resistance values.
- **Menstrual cycle in women**. Note menstrual cycle; be consistent in terms of timing for follow-up measurements.
- **Timing of measurement**. If individual is ambulatory, individual should assume a supine position for 5-10min; standardize the timing for measurements by noting the time when the individual assumes the supine position and the time when you take the measurement (eg, at 10 minutes), and ensure consistency of timing for all followup measurements. Note if individual is confined to bed.
- **Repeat measurements**. Repeat measures recommended for research studies.

The level of precision produced by both SF-BIA and MF-BIA devices is said to be good, with variability between repeat measures reported as 1%-2%. For BIS measurements, precision is more variable (i.e., 2%-3%) (Earthman, 2015).

In general, many BIA devices yield relatively valid estimates of fat-free mass and other body compartments in healthy normal-weight individuals and for large-scale epidemiological studies (i.e., mean level accuracy), but results in clinical settings, especially for individuals with fluid overload, have been mixed. None of the devices generate valid estimates for whole body composition in individuals with obesity (Earthman, 2015).

Like TOBEC (Section 14.7), BIA also depends on the differences in electrical conductivity of a weak alternating current which flows at various rates depending on the composition of the body. The current is well conducted by tissues with a high water and electrolyte content such as blood and skeletal muscle but is poorly conducted by adipose tissue, bone, and air-filled spaces which have a low water and electrolyte content. The voltage decrease of the current as it passes through the body is detected through current sensing surface electrodes, and the impedance data are recorded by the BIA device. See Kyle et al. (2004) for details of the theory of BIA.

The placement of the adhesive electrodes to the body is important and varies, depending on the device and whether whole-body or segmental measurements are sought. The standard placement is a wrist-ankle tetrapolar arrangement on the hand and foot of one side of the body. Alternatively, the electrodes can be placed on both sides of the body on both limbs using an 8-electrode arrangement, or on different segments of the body. For the stand-on BIA devices, direct contact with the electrodes is achieved at the feet and/or hands depending on the device.

The extrapolation from raw impedance data to volume depends on several key assumptions. For example, one of the assumptions is that stature is an acceptable surrogate for the unknown true length of the conductor, whereas another assumes the human body is comprised of five cylinders of uniform cross-sectional area. Unfortunately, these assumptions are often violated, especially those related to body differences in geometry in individuals with obesity and/or longer or shorter than average limbs. With the introduction of segmental and MF-BIA techniques, some of the limitations of the BIA method for these individuals have been overcome. For patients with clinical conditions in which hydration is altered (e.g., heart or liver failure), however, assuming a fixed hydration will yield body composition estimates that are inaccurate. For a detailed description of the assumptions, see Mulasi et al. (2015).

14.8.1 Single-frequency bioelectrical impedance (SF-BIA)

Single frequency BIA devices (SF-BIA) were the earliest devices developed. With these, the impedance of a weak electrical current (typically 500-800mA) and a single current frequency (i.e., typically 50kHz) is passed between two electrodes, usually located in the standard wrist-ankle tetrapolar arrangement on one side of the body, as shown in Figure 14.10, while the subject is supine in a horizontal position.

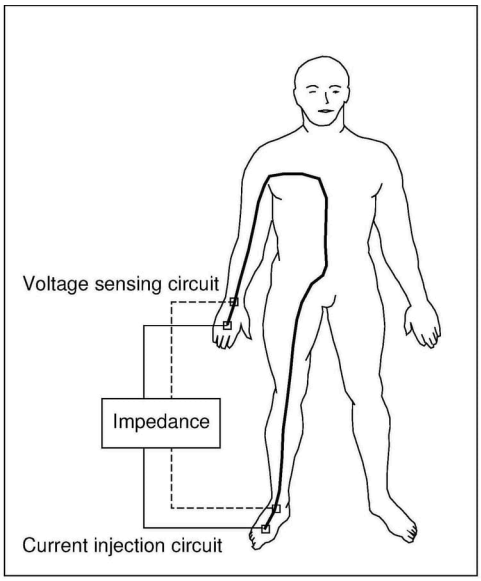


Figure 14.10. Measurement of whole-body bioelectrical impedance, showing diagrammatically the current path through the body and the positioning of the four electrodes

To derive fat-free mass, total body water (TBW) is predicted from raw impedance data by applying an appropriate regression equation and the use of the assumed hydration constant for fat-free mass (i.e., typically 0.732 for a healthy adult) (Section 14.3). However, as discussed earlier (see Section 14.3 and Section 14.4), many factors influence the hydration of fat-free mass which have the potential to limit the accuracy of the estimates of fat-free mass from SF-BIA data.

Theoretically, the impedance measured at a single frequency (e.g., 50kHz) from a SF-BIA device is unable to differentiate between extracellular (ECW) and intracellular water (ICW) (or body cell mass, BCM). Many SF-BIA devices produce predictions of these outputs based on assumptions that may hold true in healthy individuals but not for those who are obese or sick.

14.8.2 Multiple-frequency bioelectrical impedance (MF-BIA)

Multiple-frequency BIA devices (MF-BIA) measure impedance at two or more frequencies, typically at 4 or 5 frequencies, including at least one low (most commonly 5kHz) and several higher ones (typically 50, 100, 200, and 500kHz). At lower frequencies the impedance to current flow allows for the determination of the extracellular water space (ECW), whereas at the higher frequencies (i.e., > 50kHz) the impedance to current allows for the determination of ECW and intracellular water space (ICW) (i.e., TBW). The volume of ICW, derived by subtracting ECW from TBW, can be used to estimate body cell mass (BCM) based on the assumption that cells are comprised of 70% water.

Several population-specific prediction equations are available to predict ECW, TBW, ICW, and BCM based on reference values derived from appropriate criterion methods. Nevertheless, again underlying assumptions may be violated in those individuals with obesity and in certain clinical populations, leading to errors, highlighting the importance of not accepting, without question, the output from a BIA device which uses the proprietary equations of the manufacturer (Earthman, 2015).

Sex-specific BIA prediction equations, validated earlier from MF-BIA resistance measures in the U.S NHANES III in 1988-1994 (Chumlea et al., 2002), have been used to compile ratios of fat mass to fat-free mass at 5th, 50th, and 95th percentiles by sex, age group, and BMI category (i.e., underweight, normal weight, overweight, class I/II and class III obesity) for non-Hispanic persons aged 18-90y. These fat/fat-free mass reference values that account for age, sex, and BMI can be used to identify individuals at risk for body composition abnormalities (Xiao et al., 2018).

Measurements from newer MF-BIA devices can now be made across body segments such as the limb and trunk, or across a small body region such as the muscle bed of the calf. Some MF-BIA devices can be used in upright and supine positions so that they can be used for non-ambulatory or bed-bound persons.

14.8.3 Bioelectrical impedance spectroscopy (BIS)

Bioelectrical impedance spectroscopy (BIS) devices use biophysical modeling to estimate body compartments, as noted earlier; see Mulasi et al.(2015) for more details. These devices apply the electric current (typically ≤ 800μA) over a range of frequencies, from very low (e.g., 1 or 5kHz) to very high (e.g., 1000-1200kHz), measuring impedance data at ≥ 50 frequencies. This allows a more direct and individualized measure of ECW, ICW, and TBW compartments compared with those generated with SF-BIA and MF-BIA. Such measures are particularly useful in patients with altered fluid homeostasis (Mulasi et al., 2015). For those with obesity, the values for ECW and ICW can now be corrected using BMI as a surrogate for adiposity, thus allowing a more accurate assessment of body composition in individuals with obesity (Moissl et al., 2006).

Many validation studies of BIS conducted in healthy and clinical populations are available. Although good agreement at the population level can be achieved, agreement between reference methods and BIS at the individual level is poor, as noted for the measurements from SF-BIA and MF-BIA devices. This limits the assessment of whole body masses and fluid volumes in clinical settings.

14.8.4 Bioelectrical impedance vector analysis (BIVA)

Bioelectrical impedance vector analysis is a graphical procedure which uses the plot of resistance and reactance standardized for height to create a vector that can be compared with gender- and race-specific reference values from healthy population samples. BIVA can be generated from 50kHz data. It can provide information on hydration and body cell mass for patients in whom calculation of body composition is incorrect due to altered hydration (Mulasi et al., 2015). In a recent study of children with severe-acute undernutrition in Ethiopia, BIVA measurements successfully differentiated between those children who were dehydrated and those with edema (Girma et al., 2021). During treatment, edematous children lost fluid whereas non-edematous children gained small amounts of fat-free tissue. Moreover, BIVA parameters correlated with biomarkers of nutritional status (Girma et al., 2018).

14.8.5 New developments in bioelectrical impedance analysis

The questionable validity of BIA approaches for the assessment of whole body composition estimates in clinical populations, especially those with abnormal body geometry or altered fluid homeostasis, has led to the use of raw BIA measurements per se for bedside assessment of nutritional status and/or clinical outcomes. Raw BIA data are independent of the use of regression prediction models and assumptions of constant chemical composition of the fat-free mass, unlike the measurements from SF-BIA and MF-BIA devices. The raw BIA data used comprise single-frequency (50kHz) phase-sensitive measurements to determine: (i) phase angle, and (ii) the ratio of multifrequency impedance values.

Phase angle is estimated directly by a phase-sensitive BIA device without additional conversion to specific body compartments, followed by comparison with population-specific reference values. Phase angle is also independent of the body weight and height of an individual patient. The phase angle concept is based on changes in resistance and reactance as alternating currents pass through evaluated tissues, providing information on hydration status and cell mass; for more details see Lukaski et al. (2017). Low phase angles values are typically related to more-severe illnesses and worse overall health outcomes. Currently, the major challenge of using phase angle for clinical assessment is the lack of consensus on the choice of cut-points to identify malnutrition (or poor clinical outcomes) (Mulasi et al., 2015).

The ratio of impedance measured at 200kHz to impedance measured at 5kHz — termed the impedance ratio or prediction marker — is said to reflect the ratio of ECW/TBW fluid distribution, and which is currently being explored as a potential indicator of nutritional status and/or clinical outcomes. However, as stated for phase angle, the lack of a consensus on the reference cut-

points to use to identify malnutrition (or poor clinical outcomes) remains a major challenge for the use of the impedance ratio for clinical assessment (Mulasi et al., 2015). investigations to date suggest that such reference cut-points may differ depending on the population under study and the BIA device used.

Of particular interest is whether both phase angle (PA) and the impedance ratio (IR) are useful for the diagnosis of sarcopenia (with and without the presence of obesity) and malnutrition in clinical settings (Mulasi et al., 2015). Reference cut-points for both phase angle and impedance ratios by sex, ethnicity, and age-decade have been compiled from U.S NHANES 1999-2004 based on BIS data. Validation with other reference measures (e.g. DXA) is needed to assess whether PA/IR are appropriate for the assessment of nutritional status in a clinical population (Kuchnia et al., 2017).

14.8.6 Applications of bioelectrical impedance

Several population-specific regression equations have been developed based on reference methods capable of predicting not only fat-free mass, but fat mass and other compartments from 50kHz data. Examples of these reference methods include those based on a 4-component model or dual-energy X-ray absorptiometry (DXA); see Kyle et al. (2004) for more details. The regression equation chosen must be matched closely to the characteristics of the subject. However, unfortunately, many devices do not specify the equation programmed into their software so that the prediction equation used with the device may not be appropriate for the individual being measured.

Use of a fat-free mass index has also been explored for its potential to predict nutritional status and/or clinical outcomes. The index is generated from 50kHz data and is a height-corrected index of fat-free mass (i.e., fat-free mass/height squared) calculated by a standardized prediction equation and compared with reference data (Schutz et al., 2002). To date, challenges remain when applying fat-free mass index to assess nutritional status in clinical settings and more research is needed (Mulasi et al., 2015).

Interestingly, the American Society for Parenteral and Enteral Nutrition (ASPEN), does not support the use of BIA for the assessment of body composition in the clinical setting, based on their systematic review of 23 BIA studies. Their main objections included the scarcity of data on the validity of BIA in specific clinical populations, difficulties comparing studies using different BIA devices, variability in the body compartments estimated, and the proprietary nature of manufacture- specific BIA regression models to procure body composition data. For a summary and discussion of the studies reviewed which led to this conclusion, see Sheean et al. (2020).

This page titled 12.9: Bioelectrical impedance (14.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.10: Computerized tomography (14.9)

Computerized tomography (CT) is a high-resolution imaging technique that is widely used in clinical settings to quantify in vivo body composition at the tissue-organ level. Total adipose tissue, subcutaneous adipose tissue, and visceral adipose tissue can be assessed. Computerized tomography can also measure skeletal muscle, individual muscle, or muscle groups, and evaluate the quality of muscle by identifying the infiltration of fat within the muscle, a condition known as myosteatosis. Computerized tomography is especially useful for investigating quantitative and qualitative changes in muscle and fat in the trunk area where the use of DXA is limited.

Computerized tomography (CT) is based on the relationship between the attenuation of an X-ray beam and the physical density of the tissues through which the beam has passed. The known differences in attenuation of X-rays between lean soft tissue and adipose tissue can be used to distinguish these tissues as well as to determine mixtures of them. From this relationship, a two-dimensional high-resolution radiographic image of the underlying anatomy of the scan area can be constructed.

The CT scanner is made up of two components: a collimated X-ray source and detectors, and a computer that processes the scan data and produces an X-ray image. The subject lies on a movable platform within the scanner gantry. The designated area to be scanned is a plane through the middle of the central aperture of the gantry and parallel to the gantry. The X-ray beam is made to rotate around the subject, generating a cross-sectional “slice” through the patient. As the X-rays pass through the tissue, the beam undergoes attenuation, and the intensity is recorded and stored in the scanner computer. The latter then processes the stored information by using a series of complex algorithms to construct a cross-sectional image (Figure 14.11). Multiple images generated following the movement of the patient through the scanner gantry are used to produce an integrated scan of the subject.

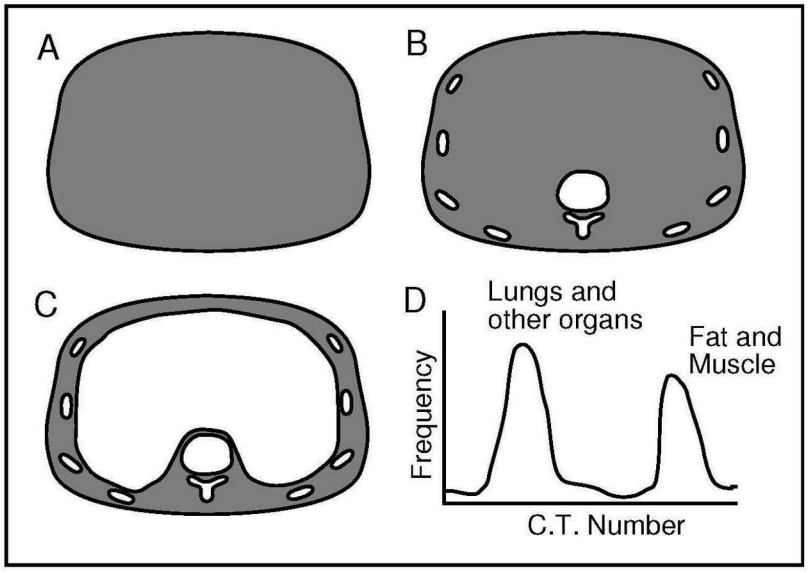


Figure 14.11. Body composition from a computerized tomography scan at the thorax level, showing three images resulting from different computer processings of the same scan. A: The air/skin interface. B: Major skeletal elements. C: Adjusted to show the interface between the lungs and other organs and the surrounding muscle and fat. D: Histogram of the pixel density from scan C.

The integrated scan shows varying degrees of shading according to the magnitude of the X-ray beam attenuation, which, in turn, depends on the physical density of the scanned tissues. Tissues with a greater density cause a greater absorption of X-ray energy

and, consequently, a higher attenuation value. The demarcation between tissues of differing density — adipose tissue, skeletal muscle, bone, and visceral organs — can be very good.

The use of computerized tomography is limited by its high-dose radiation exposure and high cost. Efforts have been made to reduce the radiation dose by using a lower-dose protocol, essential when planning a CT scan in a child (Sorantin et al., 2013). This protocol has little deterioration in image quality (Kapur et al., 2021). Single-slice scanning can also be used to reduce the radiation dose.

A new technique with extremely low radiation exposure, shorter scan time, and relatively low cost compared to whole body CT, termed “peripheral quantitative computer tomography” (pQCT) has been developed. This new technique has been used primarily to investigate bone mineral content and to assess bone fragility in older patients (Jiang et al., 2018). However, despite these efforts, radiation exposure is still high, restricting the use of CT images for clinical use during disease treatment rather than for body composition assessment.

14.9.1 Application of computerized tomography

The method has several uses. It can be used to assess changes in the visceral organ mass in undernutrition and obesity, to portray the distribution of subcutaneous versus visceral adipose tissue, and to establish bone density in osteopenia (Heymsfield et al., 1987). New techniques with improved spatial resolution can be used to detect and measure fat in areas of the body where fat is not physiologically stored, such as the liver, pancreas, heart, and skeletal muscle. The fat deposited in these areas is termed visceral ectopic fat and is known to contribute to the development of coronary artery disease as well as other cardiovascular disorders (Neeland et al., 2019).

The region of special interest for the study of body composition using CT scans is the third lumbar vertebra (L3). This is the region where a single image of a cross- sectional area provides the best correlation of whole-body skeletal muscle volume. The validity and accuracy of body composition measurements by CT based on cross-sectional area of tissues have been validated by studies of human cadavers, although no validation of the calculation of tissue volume has been performed (Fosbøl & Zerahn, 2015).

Increasingly, CT scans are also being used to identify CT-defined sarcopenia, a condition associated with a decrease in muscle mass and function which, although originally described in the elderly, is also of concern among the chronically ill nonelderly (Peterson & Braunschweig, 2016). To date, however, there are no clear diagnostic cutoff values for CT to identify sarcopenia based on skeletal muscle mass, which has limited the application of CT for clinical use.

This page titled 12.10: Computerized tomography (14.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.11: Magnetic resonance imaging (14.10)

Unlike computer tomography or DXA, magnetic resonance imaging (MRI) does not use ionizing radiation so that the technique can be used on infants, and for long-term follow-up when multiple scans on the same person are required. The MRI technique is mainly used to evaluate the quantity and distribution of adipose tissues and skeletal muscle mass, although it can also detect changes in body composition, even in the presence of small body weight changes (Lemos & Gallagher, 2017). In older equipment, scanning times of 10min were necessary, but in more modern equipment, this has been reduced to under 2min. Nevertheless, the MRI technique requires technical expertise, is expensive, and the equipment is bulky (Prada & Heymsfield, 2014).

Magnetic resonance imaging uses differences in the nuclear magnetic resonance properties of hydrogen atoms in organic and non-organic environments to distinguish signals originating from fat, fat-free mass, and free water. The hydrogen protons behave slightly differently in adipose versus lean tissues. The differences are in the relaxation time that it takes for the nuclei to release the radio-frequency-induced energy and return to a random configuration. These differences can be used to map the distribution of adipose versus lean tissue in the body (Ross, 1996).

The imaging process involves placing the subject in a very strong magnetic field. Some of the nuclei in the body attempt to align themselves relative to the applied field. The effect is particularly marked for ^{1}H protons. Only a small fraction of the protons become aligned, but they are sufficiently numerous for the effect to be detectable when the field is removed or altered. It is then that the differences between the lean and adipose tissue become apparent.

This page titled 12.11: Magnetic resonance imaging (14.10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.12: Dual energy X-ray absorptiometry (14.11)

Dual-energy X-ray absorptiometry (DXA) is a technique that can be used with individuals (except pregnant women) across the entire age range and at relatively low cost. DXA is widely used to assess body composition, with an overall precision that exceeds that of any other body composition method(Prada & Heymsfield, 2014).

With the latest generation of densitometers, body composition can be assessed with a single whole-body scan so that radiation exposure is low with a minimal acquisition time. In addition, the newer instruments enable individuals with extreme obesity to be scanned.

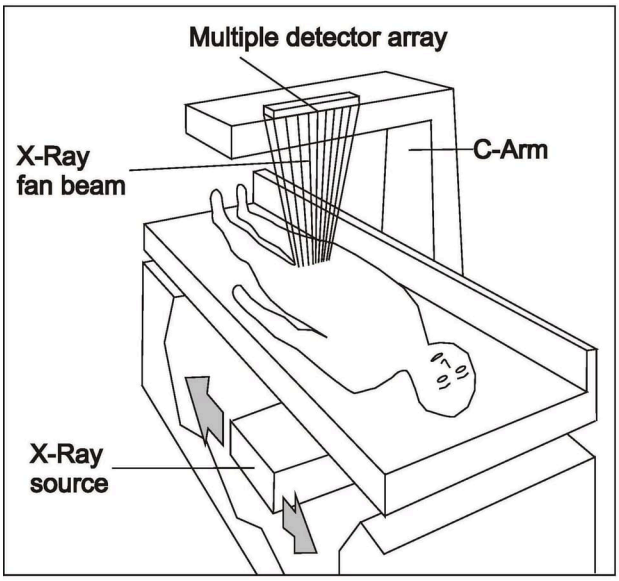


Figure 14.12. Dual-energy X-ray scanner with multiple detector array and X-ray fan beam.

Modern DXA scanners use an X-ray source, a detector, and an interface with a computer system for imaging the scanned area of interest. The source generates X-rays at two different energy levels — 40KeV and 70-100KeV. These pass through the body and are identified by the photon detector that measures the amount of photon energy absorbed at the two energy levels by the soft tissue and bone. Low-density material (i.e., soft tissue) allows more photons to pass through, so they attenuate the X-ray beam less than materials with a higher density, such as bone. An image is generated as the photon detector measures the differential attenuation (or absorption) of the low and high X-rays by the soft tissue and bone. Only in regions of the body where no bone is present can measurements of the relative proportions of fat and lean tissues be made; see Prada & Heymsfield (2014) for more details.

During an investigation, the subject, on a movable bed, is scanned rectilinearly from head to toe. The X-ray source is situated beneath the scanning bed, and scans usually take from 5-20min depending on type of scan and software selected (Figure 14.12).

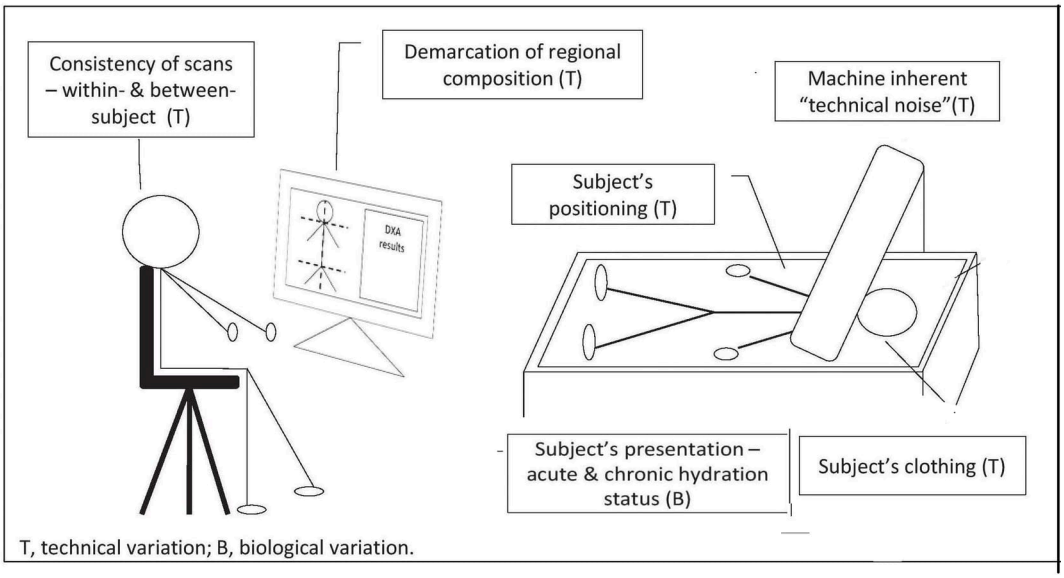


Figure 14.13. Sources of variability or error when carrying out a DXA scan to measure body composition. Modified from Nana et al. (2015).

The accuracy of DXA results is influenced by both technical and biological factors; these are summarized in Figure 14.13. The technical factors include differences in machine calibration procedures, the protocols used in demarcating the regions (arms, legs and trunk) in estimates of body composition, the positioning of the subject on the scanning bed, and the effect of clothing. Sources of biological variability that can affect accuracy include hydration status because the hydration of fat-free mass is assumed to be constant (i.e., about 73%). Hence, when changes in hydration status are large (i.e., higher than 5%), as may occur in the presence of diseases with water retention (e.g., heart, kidney or liver failure), the fat-free mass is over-estimated (Nana et al., 2015). In older DXA models, obesity was also a factor limiting accuracy, but newer instruments allow larger and heavier individuals to be scanned. Finally, the ability to compare body composition estimates between DXA manufacturers is limited because of variations in the hardware and software packages used by manufacturers, highlighting the urgent need for universal calibration of DXA machines (Prada & Heymsfield, 2014).

Several investigators have tried to establish the absolute accuracy of gross measurements of body composition in vivo performed using DXA. Many of these studies have examined correlations between measurements made by DXA on pig cadavers and subsequent direct chemical analysis; results have been mixed (Jebb, 1997).

14.11.1 Applications of DXA

Dual-energy X-ray absorptiometryis now the primary method for generating accurate data on bone mineral content and density for the diagnosis of osteoporosis (see Chapter 23 for more details). DXA is also widely used to assess body composition, with an overall precision that exceeds that of any other body composition method (Prada & Heymsfield, 2014). Both total and regional estimates of the three body compartments — fat-free mass, body fat, and bone mineral content — can be determined. However, DXA does not have the ability to discriminate between visceral, subcutaneous, and ectopic fat present in organs such as liver or muscle.

Estimates of the appendicular skeletal muscle mass (ASM) can be obtained from DXA. These estimates are generated by summing the amount of lean soft tissue (LST) of the arms and legs, as they are composed mainly of muscle (except for a small amount of connective tissue and skin) (Prada & Heymsfield, 2014). From these estimates, the ASM index can be calculated by dividing the arm and leg lean soft tissue (LST) by height squared. Both the ASM and the ASM index values are considered the gold standard method for the diagnosis of sarcopenia. For published cutoff points to define sarcopenia based on DXA, see Earthman (2015).

In a systematic review based on eight DXA studies by the American Society for Parenteral and Enteral Nutrition (ASPEN), DXA was recommended for the assessment of fat mass in patients with a variety of disease states, but not for fat-free mass (Sheean

et al., 2020). They stressed that the validity of DXA for assessment of fat-free mass in any clinical populations remains unknown. Clearly, additional research is needed to evaluate the validity of DXA for quantifying fat-free mass in clinical populations.

This page titled 12.12: Dual energy X-ray absorptiometry (14.11) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.13: Ultrasound (14.12)

Diagnostic ultrasound, also called sonography, is an imaging technique used in clinical settings to measure various tissue thicknesses, including muscle, bone, and subcutaneous and visceral adipose tissue. The device is portable, low cost, and capable of making fast and noninvasive regional estimates of body composition with no exposure to ionizing radiation. Consequently, this scanning technique can be used in children and pregnant women and is suitable for large epidemiological studies, although the results generated are highly dependent on the skill of the operator.

The ultrasound technique involves the use of high-frequency sound waves from a transducer. These penetrate the skin surface and pass through the adipose tissue until they reach the muscle tissue. At the adipose-muscle tissue interface, a proportion of the sound waves are reflected back to the transducer as echoes. Hence, the transducer both transmits and receives the ultrasound. The degree of reflection is dependent on the changes in acoustic impedance (i.e., the product of tissue density and acoustic velocity) between two tissue interfaces. The higher the acoustic impedance, the stronger the generated reflection, and thus, the better the quality of the image.

The acoustic impedance of fat and muscle are somewhat similar and much lower than bone which has a relatively high acoustic impedance; air has almost no impedance. Consequently, there is a weaker echo at the fat-muscle interface than at the muscle-bone interface. When the transducer receives the beam, it converts the echo into electric signals to form a two-dimensional image. The relative strength, or amplitude, of echoes is apparent from the brightness of the image on the computer screen, with strong reflections appearing white, weaker reflections grey, and no echoes black. As a result, a grey-scale image is produced with white borders for the skin-subcutaneous fat and muscle-bone interfaces and a visible, but less distinct border for the fat-muscle interface; see Wagner (2013) for more details.

However, the interpretation of ultrasound images is difficult and depends on the technical expertise of the operator who must identify the interfaces and measure the thickness of the tissue layer of interest using electronic calipers. Care is needed to identify and place the two caliper points at the boundaries of the tissue to be measured. An automated discrimination method is available to identify the tissue boundaries for some ultrasound devices, although studies comparing the automated method with the manual discrimination are limited (Wagner, 2013).

To use ultrasound, the measurement site is marked with a water-soluble transmission gel that provides acoustic contact without depression of the dermal surface. The high-resolution transducer is then placed without loss of contact with the skin so that the ultrasonic beam is perpendicular to the tissue interfaces at the marked site. A transducer receives the echoes and translates them into depth readings viewed on a computer screen. Thicknesses of 100mm or more can be measured, and density interfaces can be detected with an accuracy of 1mm. The tissue is not compressed, thereby eliminating errors associated with variations in the compressibility of skinfolds (Fanelli & Kuczmarski, 1984).

14.12.1 Applications of Ultrasound

Visualization of a fetus during a prenatal examination is the most familiar application of ultrasound. However, ultrasound is also used to measure the quality and quantity of skeletal muscle mass, an index of lean soft tissue, in the elderly, patients with cystic fibrosis, and patients confined to bed rest (Prada & Heymsfield, 2014). Alternatively, measurement of the thickness of subcutaneous fat can be used to monitor changes in body composition of hospital patients receiving nutritional support.

In general, the ultrasound method provides a reasonable estimate of adipose tissue thickness in humans, compared to total body electrical conductivity (TOBEC) and skinfold caliper techniques (Fanelli & Kuczmarski, 1984). For obese persons especially, ultrasound may be superior to skinfold caliper techniques for measuring subcutaneous fat (Kuczmarski et al., 1987). The method can also be used to measure thickness of other tissues such as muscle and bone, as noted earlier; for more details see Mayans et al. (2012) and Karjalainen et al. (2008).

An ultrasound system has also been designed specifically for body composition which could be the user-friendly ultrasound alternative to skinfolds and other field methods for estimat;ing percentage body fat (Wagner, 2013). However, currently the diverse technology across the commercially available devices and the lack of standardized measurement protocols makes it difficult to compare results across studies (Lee et al., 2019). The American Society for Parenteral and Enteral Nutrition (ASPEN) did not recommend the use of ultrasound to assess body composition in a clinical setting, based on their systematic review of seven studies (Sheean et al., 2021). More research is needed to develop consensus on the optimal method to conduct ultrasound

measurements and to generate population-specific reference data so that the method can be used to assess lean tissue and diagnose malnutrition in a clinical setting (Earthman, 2015).

This page titled 12.13: Ultrasound (14.12) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

12.14: Summary - Method comparisons (14.13)

Increasingly, clinicians are using the more robust and reproducible in vivo methods described in this chapter to measure body composition in view of the limitations of BMI as a proxy for adiposity and the recognition that differences in body composition are associated with an increased risk of diseases across the lifespan. Such diseases may include AIDS-associated wasting, diabetes, osteoporosis, sarcopenia, cardiovascular disease, and anorexia nervosa. Even during early infancy, changes in body composition (both fat and fat-free mass) can provide more understanding of the initiators and mediators of the developmental origins of adult cardiometabolic disease (McMillen et al., 2005).

However, no single in vivo method has the ability to track body composition accurately from infancy to adulthood; all have strengths and technological limitations. Fields and co-workers (2015) have compared selected in vivo methods for assessing whole-body composition across the life span; details are shown in Table 14.6.

Table 14.6. Comparison of selected methods for determining whole-body composition across the life span.
C, cost limits widespread use; D, difficult to perform but still possible; I, impractical (maybe impossible) due to age- and compliance probative data support its validity, precision, and reliability in this population; PL, possible—limitations exist in interpretation to whol boards may limit its use; UE, may be unethical due to radiation; VA, validity is ambiguous or weak in this population. [a]gap exists betw

Method	Infancy (< 6mo)	Childhood (2-6y)
Hydrostatic weighing	NA	I
Air displacement plethysmography	P	Pa|D
Whole-body counting for K	UE|C	UE|C
Dual-energy X-ray absorptiometry	D|R	D|R
Bioelectrical impedance	VA|NV	VA|NV
Computed tomography	UE	UE
Isotope dilution	I	D|P
Magnetic resonance	D|C	D|I|C
Ultrasound	NV	NV

When compiling this table, the investigators considered the technological limitations of each method, and the following four criteria:

- Cost of the measurement / device
- Level of patient cooperation needed for accurate results
- Feasibility in obtaining an interpretable result given the patient's age
- Appreciation of the age of the patient relative to the assumption of the underlying principles in the chosen method.

Based on their review of the methods listed for whole-body assessment in Table 14.6, Fields et al. (2015) chose air displacement plethysmography (Section 14.6.3) as the method with the highest degree of accuracy and reliability and with the least degree of technical error across the life span. Currently a "suite" of air displacement plethysmographic devices are available which can be used to track and monitor body composition from birth, into adolescence and throughout adulthood; they are described in Section 14.6.3 and have also been reviewed by Fields et al. (2015).

This page titled 12.14: Summary - Method comparisons (14.13) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13: BIOMARKERS (CHAPTER 15)

Abstract

Nutritional biomarkers are defined as biological characteristics that can be objectively measured and evaluated as indicators of normal biological or pathogenic processes, or as responses to nutrition interventions. They can be classified as: (i) biomarkers of exposure; (ii) biomarkers of status; and (iii) biomarkers of function. Biomarkers of exposure are intended to measure intakes of foods or nutrients using traditional dietary assessment methods or objective dietary biomarkers. Status biomarkers measure a nutrient in biological fluids or tissues, or in the urinary excretion of a nutrient or its metabolites; these ideally reflect total body nutrient content or the status of the tissue store most sensitive to nutrient depletion. Functional biomarkers, subdivided into biochemical and physiological or behavioral biomarkers, assess the functional consequences of a nutrient deficiency or excess. They may measure the activity of a nutrient-dependent enzyme or the presence of abnormal metabolic products in urine or blood arising from reduced activity of the enzyme; these serve as early biomarkers of subclinical deficiencies. Alterations in DNA damage, in gene expression and in immune function are also emerging as promising functional biochemical biomarkers. Disturbances in functional physiological and behavioral biomarkers can occur with more severe nutrient deficiencies, often involving impairments in growth, vision, motor development, cognition, in response to vaccination, and the onset of, or an increase in, depression. Such functional biomarkers, however, lack both sensitivity and specificity as they are often also affected by social and environmental factors. Outlined here are the principles and procedures that influence the choice of the three classes of biomarkers, as well as confounding factors that may affect their interpretation. A brief review of biomarkers based on new technologies such as metabolomics, etc., is also provided. Methods for evaluating biomarkers at the population and individual level are also presented.

13.1: Biomarkers to assess nutritional status (15.1)

13.2: Biomarkers of exposure (15.2)

13.3: Biomarkers of status (15.3)

13.4: Biomarkers of function (15.4)

13.5: Factors affecting choice of nutritional biomarkers (15.5)

13.6: Evaluation of the selected nutritional biomarker (15.6)

13.7: Application of new technologies (15.7)

This page titled 13: Biomarkers (Chapter 15) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13.1: Biomarkers to assess nutritional status (15.1)

Nutritional biomarkers are increasingly important with the growing efforts to provide evidence-based clinical guidance, and advice on the role of food and nutrition in supporting health and preventing disease. A nutritional biomarker has been defined by the Biomarkers of Nutrition and Development (BOND) program as a biological characteristic that can be objectively measured and evaluated as an indicator of normal biological or pathogenic processes, and/or as an indicator of responses to nutrition interventions (Raiten and Combs, 2015). Thus nutritional biomarkers can be measurements based on biological tissues and fluids, on physiological or behavioral functions, and more recently, on metabolic and genetic data that in turn influence health, well-being and risk of disease. Most useful are nutritional biomarkers that distinguish deficiency, adequacy and toxicity, and which assess aspects of physiological function and/or current or future health. Increasingly, understanding the effect of diet on health requires the study of mechanisms, not only of nutrients but also of other bioactive food constituents at the molecular level. Hence, there is also a need for molecular biomarkers that allow the detection of the onset of disease, in, ideally, the pre-disease state. Unfortunately, nutritional biomarkers are often affected by technical and biological factors other than changes in nutritional status, which can confound the interpretation of the results.

Nutritional biomarkers are used to support a range of applications at both the population and individual level; these applications are listed below.

At the population level

- National Nutrition surveys: assess overall nutritional status of populations
- Nutrition Screening: identify persons “at risk” in the population via cut-offs
- Surveillance: continuous monitoring of nutritional status of selected population groups over time (e.g., U.S. NHANES and U.K. Diet and Nutrition Survey Rolling Program)
- Monitoring and Evaluation: monitor coverage of / compliance with nutrition policies; evaluate the efficacy and/or effectiveness of public health programs and interventions over time; substantiate health claims.

At the individual level

- In apparently healthy patients: assess reserves, pool size, tissue amounts of the nutrient; determine response to clinical treatment of a nutrient deficiency or disease state
- In “sick” patients: determine status for a specific clinical problem; reflect current status of deficiency or clinical disease; predict future risk of disease or long-term functional outcome if abnormal values persist.

Application list modified from Raiten et al. (2011).

15.1.1 Classification of biomarkers

BOND has classified nutritional biomarkers into three groups, shown in Box 15.1, based on the assumption that an intake-response relationship exists between the biomarker of exposure (i.e., nutrient intake) and the biomarkers of status and function. Nevertheless, it is recognized that a single biomarker may not reflect exclusively the nutritional status of that single nutrient, but instead be reflective of several nutrients, their interactions, and metabolism. In addition, a nutritional biomarker may not be equally useful across different applications or life-stage groups where the critical function of the nutrient or the risk of disease may be different.

Biomarkers of exposure are intended to assess what has been consumed, and, where possible, take into account bioavailability, defined as the proportion of the ingested nutrient that is absorbed and utilized through normal metabolic pathways (Hurrell et al., 2004). Biomarkers of exposure can be based on measurements of nutrient intake obtained using traditional dietary assessment methods. Alternatively, depending on the nutrient, nutrient exposure can be measured indirectly, based on surrogate indicators termed “dietary biomarkers”. These are intended to provide a more objective measure of dietary exposure that is independent of the measurement of food intake.

Box 15.1. Classification of nutritional biomarkers

- Biomarkers of “exposure”: food or nutrient intakes; dietary patterns; supplement usage. Assessed by:

- Traditional dietary assessment methods
- Dietary biomarkers: indirect measures of nutrient exposure

- Biomarkers of “status”: body fluids (serum, erythrocytes, leucocytes, urine, breast milk); tissues (hair, nails)
- Biomarkers of “function": measure the extent of the functional consequences of a nutrient deficiency: serve as early biomarkers of subclinical deficiencies.
  - Functional biochemical: enzyme stimulation assays; abnormal metabolites; DNA damage
  - Functional physiological/behavioral”: more directly related to health status or disease such as vision, growth, immune function, taste acuity, cognition, depression. These biomarkers impact on clinical and health outcomes.



Biomarkers of status measure either a nutrient in biological fluids or in tissues, or the urinary excretion rate of the nutrient or its metabolites, often with the aim of assessing where an individual or population stands relative to an accepted cut-off (e.g., adequate, marginal, deficient). Ideally, the biomarker selected should reflect either the total body content of the nutrient or the size of the tissue store that is most sensitive to depletion. In practice, such biomarkers are not available for many nutrients. Furthermore, even if levels of the nutrient or metabolite in the biological tissue or fluid are “low”, they may not necessarily reflect the presence of a pathological lesion. Alternatively, their significance to health may be unknown.

Biomarkers of function are intended to measure the extent of the functional consequences of a specific nutrient deficiency or excess, and hence have greater biological significance than the static biomarkers. Increasingly, functional biomarkers are also being used as substitutes for chronic disease outcomes in studies of associations between diet and chronic disease. When used in this way, they are termed “surrogate biomarkers”; see Yetley et al. (2017) for more details. Functional biomarkers can be subdivided into two groups: functional biochemical, and functional physiological or behavioral, biomarkers. In some cases functional biochemical biomarkers may serve as early biomarkers of subclinical deficiencies by measuring changes associated with the first limiting biochemical system, which in turn affects health and well-being. They may involve the measurement of an abnormal metabolic product in urine or blood or the activity of a nutrient-dependent enzyme. Alterations in DNA damage, in gene expression and in immune function are also emerging as promising functional biochemical biomarkers, some of which may become accepted as surrogate biomarkers for chronic disease.

Functional physiological and behavioral biomarkers are more directly related to health status and disease than are the functional biochemical biomarkers. Disturbances in these biomarkers are generally associated with more prolonged and severe nutrient deficiency states, or risk of chronic diseases. Examples include measurements of impairment in growth, of response to vaccination (as a biomarker of immune function), of vision, of motor development, cognition, depression, and high blood pressure, all of which are less invasive and easier to perform than many biochemical tests. However, these functional physiological and behavioral biomarkers often measure the net effects of contextual factors that may include social and environmental factors as well as nutrition, and hence lack sensitivity and specificity as nutrient biomarkers (Raiten and Combs, 2015), or as surrogate biomarkers substituting for clinical endpoints (Yetley et al., 2017).

15.1.2 Factors that may confound the interpretation of nutritional biomarkers

Unfortunately, nutritional biomarkers are affected by several factors, other than the effects of a change in nutritional status, which may confound their interpretation. These factors may include technical issues related to the quality of the specimens and their analysis, participant and health-related characteristics, and biological factors. These factors are listed in Box 15.2. Knowledge of

their effects on the biomarkers for specific nutrients is discussed more fully in the nutrient-specific chapters.

Box 15.2. Technical, health, biological and other factors which may confound the interpretation of nutritional biomarkers

- Analytical quality performance factors: accuracy, precision, sensitivity, specificity, validity, predictive value
- Specimen-related limitations: half-life (days, weeks, months, years); stability (quality of the sample); biological variation; storage; contamination; hemolysis
- Participant-related factors: age, sex, ethnicity, race, genetic predisposition to disease, physiological state, hormonal status, supplement use, physical activity level, lifestyle, environment, recent dietary intake
- Biological factors: homeostatic regulation, diurnal and/or circadian variation, fasting status; nutrient interactions
- Health-related factors: medication use, inherited or acquired diseases, inflammation, stress; environmental enteropathy; obesity; unusual weight loss

The influence of these factors (if any) on each biomarker should be established before carrying out the tests, because these confounding effects can often be minimized or eliminated (Box 15.3). For example, in nutrition surveys the effects of diurnal variation on the concentration of nutrients such as zinc and iron in plasma can be eliminated by collecting the blood samples from all participants at a standardized time of the day. When factors such as age, sex, race, and physiological state influence the biomarker, the observations can be classified according to these variables. The influence of drugs, hormonal status, physical activity, weight loss, and the presence of disease conditions on the biomarker, can also be considered if the appropriate questions are included in a questionnaire.

Box 15.3. Strategies to overcome the effects of confounders on nutritional biomarkers

- Use standardized methods to collect, process, and analyze
- Classify observations by life-stage/sex/ethnicity
- Record medications, supplements; hormonal status; physical activity; obesity; health status, disease
- Avoid using cutoffs mismatched for assay
- Assess Hb variants and malaria, where appropriate
- Adjust for intra-individual variation with replicate measures
- Measure CRP & AGP; apply BRINDA correction to adjust for inflammation where necessary
- Measure multi-micronutrient biomarkers where co-existing deficiencies exist
- Combine biomarkers instead of using only one to enhance specificity

During an infectious illness, after physical trauma, with inflammatory disorders, and with obesity and diabetes, certain systemic changes occur, referred to as the “acute-phase response”, to prevent damage to the tissues by removing harmful molecules and pathogens. The local reaction is inflammation. During this reaction, circulating levels for certain micronutrient biomarkers — for example, zinc, iron, copper, and vitamin A — are altered, often due to a redistribution in body compartments, but these changes do not correspond to changes in micronutrient status. Hence, systemic changes due to the acute phase response must be assessed together with micronutrient biomarkers to ensure a more reliable and valid interpretation of the micronutrient status assessment at both the individual and population levels. Such systemic changes can be detected by measurement of elevated concentrations of several plasma proteins, of which C-reactive protein (CRP) and α-1-acid glycoprotein are recommended (Raiten et al., 2015).

This page titled 13.1: Biomarkers to assess nutritional status (15.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13.2: Biomarkers of exposure (15.2)

Biomarkers of exposure can be based on direct measurements of nutrient intake using traditional dietary assessment methods, or indirect measurements using surrogate indicators termed “dietary biomarkers”.

Traditional dietary assessment methods include 24h recalls, food records and food frequency questionnaires, the choice depending primarily on the study objectives, the characteristics of the respondents, the respondent burden, and the available resources. Each method has its own strengths and limitations; see Chapter 3 for more details. For all dietary methods, care must be taken to ensure that information on any use of dietary supplements and/or fortified foods is also collected. Seasonality must also be taken into account where necessary (e.g., for vitamin A intakes). In the absence of appropriate food composition data for the nutrient of interest, duplicate diet composites can be collected for chemical analysis.

Nutrient intakes calculated from food composition data or determined from chemical analysis of duplicate diet composites represent the maximum amount of nutrients available and do not take into account bioavailability. The bioavailability of nutrients can be influenced by several dietary and host-related factors; see Gibson (2007) for a detailed discussion of these factors. Unfortunately, factors affecting the bioavailability of many nutrients are not well understood, with the exception of iron and zinc. Algorithms have been developed to estimate iron and zinc bioavailability from whole diets and are described in Lynch et al. (2018) and the International Zinc Nutrition consultative Group (IZiNCG) Technical Brief No. 03 (2019). Alternatively, qualitative systems that classify diets into broad categories of iron (FAO/WHO, 2002) and zinc (FAO/WHO, 2004) bioavailability based on various dietary patterns can be used.

Given the challenges with the traditional dietary methods, there is increasing interest in the use of dietary biomarkers as objective indicators of dietary exposure. Dietary biomarkers can be classified into three groups: recovery, concentration, and predictive — each has distinctive properties, as shown in Box 15.4. Several criteria must be considered when selecting a dietary biomarker. These include the half-life of the biomarker, day-to-day intra- and inter-individual variability, the requirements for sample collection, transport, storage and analysis, and the impact of potential biological confounders that may cause variation in biomarker concentrations, unrelated to the level of the dietary component of interest.

Examples for each of the three groups of dietary biomarkers are shown in Box 15.4. In general, nutrient levels in fluids such as urine and serum tend to reflect short-term (i.e., recent) dietary exposure, those in erythrocytes are medium-term (e.g., for fatty acids; folate), whereas examples of long-term biomarkers are nutrient levels in adipose tissue (for fatty acids), toenails or fingernails (for selenium), and scalp hair samples (for chromium). In some circumstances, the time integration of exposure of the urinary dietary biomarkers can be enhanced by obtaining urine samples at several points in time. For more specific details of nutrient levels in urine as dietary biomarkers, see Section 15.3.12.

Box 15.4. Classification and properties of dietary biomarkers

- **Recovery biomarkers**
  - Measure total excretion of marker over a defined time period
  - Excretion is a fixed proportion of intake with only negligible inter-individual variation.
  - Best suited to measure absolute intake
  - Examples include: urinary N_2 for protein, K and Na in 24hr urines; doubly-labeled water for short-term energy expenditure
- **Concentration biomarkers**
  - Based solely on the concentration of the biomarker
  - Provide no information on physiological balance and excretion
  - Cannot be translated into absolute levels of intake
  - Positively correlated with intake, so can be used for ranking
  - Examples include: total carotenoids (fruit + vegetable intake): plasma vitamin C; phenols
- **Predictive biomarkers**
  - Incomplete recovery
  - Stable and time-dependent, and a high correlation with intake
  - Used in medicine to predict who is likely to respond to therapy
  - Rank between concentration and recovery biomarkers in terms of ability to estimate absolute intake

- Examples include: urinary sucrose and fructose for sugar intake

Modified from Kuhnle (2012).

Research on nutritional biomarkers for assessing the intake of specific foods, food groups, or combinations that describe food patterns rather than nutrients *per se,* is also emerging in an effort to improve the assessment of the relationships between diet, functional outcomes, and chronic disease. Examples include urinary excretion of proline betaine as a biomarker of citrus fruit; 1-methylhistidine and 3-methylhistidine as biomarkers of meat consumption; sucrose and fructose as predictive biomarkers of sugar intake; alkylresorcinol (in urine and plasma) as a possible whole grain wheat / rye biomarker; and plasma phospholipid pentadecanoic acid as a biomarker of dairy consumption (Hedrick et al., 2012). Use of the abundance of ^{13}C (a stable isotope of Carbon) in finger stick blood samples is also being investigated as a biomarker for self-reported intakes of cane sugar and high fructose corn syrup (Hedrick et al., 2016; MacDougall et al., 2018). More research is required to better understand, interpret, and validate the existing dietary biomarkers, as well as to develop and validate new ones.

This page titled 13.2: Biomarkers of exposure (15.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13.3: Biomarkers of status (15.3)

Biomarkers based on nutrients in biological fluids and tissues are frequently used as biomarkers of status, and in some cases, of exposure. Measurements of (a) concentrations of a nutrient in biological fluids or tissues, or (b) the urinary excretion rate of a nutrient or its metabolite can be used. The biopsy material most frequently used for these biomarkers is whole blood or some fraction of blood. Other body fluids and tissues, less widely used, include urine, saliva, adipose tissue, breast milk, semen, amniotic fluid, hair, toenails, skin, and buccal mucosa. Four stages are involved in the analysis of these biopsy materials: sampling, storage, preparation, and analysis. Care must be taken to ensure that the appropriate safety precautions are taken at each stage. Contamination is a major problem for trace elements, and must be controlled at each stage of their analyses, especially when the expected analyte levels are at or below concentrations of 1×10^{-9}g.

Ideally, as discussed above, the nutrient content of the biopsy material should reflect the level of the nutrient in the tissue most sensitive to a deficiency, and any reduction in nutrient content should reflect the presence of a metabolic lesion. In some cases, however, the level of the nutrient in the biological fluid or tissue may appear adequate, but a deficiency state still arises: homeostatic mechanisms maintain concentrations within the biological specimen, even when intakes are marginal or inadequate (e.g., serum calcium, retinol or serum zinc). Alternatively, a metabolic defect may prevent the utilization of the nutrient.

15.3.1 Blood

Samples of blood are readily accessible, relatively noninvasive, and generally easily analyzed. They must be collected and handled under controlled, standardized conditions to ensure accurate and precise analytical results. Factors such as fasting, fluctuations resulting from diurnal variation and meal consumption, hydration status, use of oral contraceptive agents or hormone replacement therapy, medications, infection, inflammation, stress, body weight and genotype are among the many factors that may confound interpretation of the results (Hambidge, 2003; Potischman, 2003; Bresnahan and Tanumihardjo, 2014).

Serum / plasma carries newly absorbed nutrients and those being transported to the tissues and thus tends to reflect recent dietary intake. Therefore, serum / plasma nutrient levels provide an acute, rather than long-term, biomarker of nutrient exposure and/or status. The magnitude of the effect of recent dietary intake on serum / plasma nutrient concentrations is dependent on the nutrient, and where necessary, can be reduced by collecting fasting blood samples. Alternatively, if this is not possible, the time interval since the preceding meal can be recorded, and incorporated into the statistical analysis and interpretation of the results (Arsenault et al., 2011).

For those nutrients for which concentrations in serum / plasma are strongly homeostatically regulated, concentrations in serum / plasma may be near-normal (e.g., calcium, zinc, vitamin A, Figure 15.1), even when there is evidence of functional impairment (Hambidge, 2003). In such cases, alternative biomarkers may be needed.

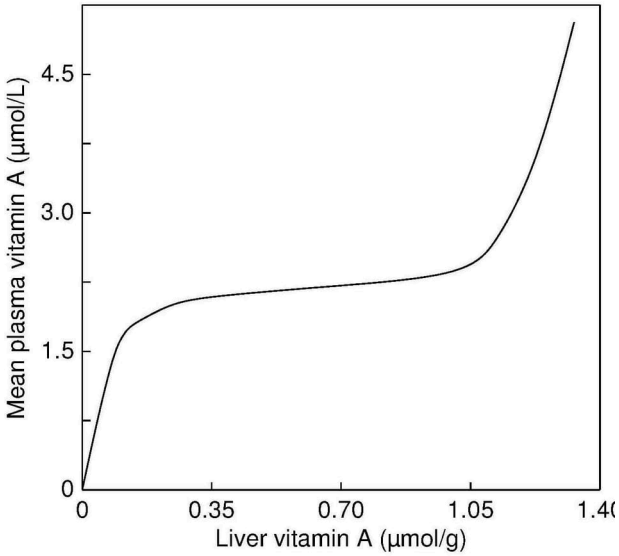


Figure 15.1. Hypothetical relationship between mean plasma vitamin A levels and liver vitamin A concentrations. From: Olson, 1984, . with permission of Oxford University Press.

The risk of contamination during sample collection, storage, preparation, and analysis is a particular problem in trace element analysis of blood. Trace elements are present in low concentrations in blood but are ubiquitous in the environment. Details of strategies to reduce the risk of adventitious sources of trace-element contamination are available in the International Zinc Nutrition Consultative Group (IZiNCG) Technical Briefs (2007, 2012). In addition, for certain vitamins such as retinol and folate, exposure to bright light and high temperature should be avoided, and for serum folate, suitable antioxidants (e.g., ascorbic acid, 0.5% w/v) are added to samples to stabilize the vitamin during collection and storage(Bailey et al., 2015; Tanumihardjo et al., 2016).

Additional confounding factors in the collection and analysis of micronutrients in blood are venous occlusion, hemolysis (IZiNCG Technical Brief No.6, 2018), use of an inappropriate anticoagulant, collection-separation time, leaching of divalent cations from rubber stoppers in the blood collection tubes, and element losses produced by adsorption on the container surfaces or by volatilization during storage (Tamura et al., 1994; Bowen and Remaley, 2013). For trace element analysis, trace-element-free evacuated tubes with siliconized rather than rubber stoppers must be used.

Serum is often preferred for trace element analysis because, unlike plasma, risk of adventitious contamination from anti-coagulants is avoided, as is the tendency to form an insoluble protein precipitate during freezing. Nevertheless, serum is more prone than plasma to both contamination from platelets and to hemolysis. For capillary blood samples, the use of polyethylene serum separators with polyethylene stoppers are recommended for analysis of trace-elements (King et al., 2015).

15.3.2 Erythrocytes

The nutrient content of erythrocytes reflects chronic nutrient status because the lifespan of these cells is quite long (≈ 120d). An additional advantage is that nutrient concentrations in erythrocytes are not subject to the transient variations that can affect plasma. The anticoagulant used for the collection of erythrocytes must be chosen with care to ensure that it does not induce any leakage of ions from the red blood cells. At present, the best choice for trace element analysis is heparin (Vitoux et al., 1999).

The separation, washing and analysis of erythrocytes is technically difficult, and must be carried out with care. For example, the centrifugation speed must be high enough to remove the extracellular water but low enough to avoid hemolysis. Care must be taken to carefully discard the buffy coat containing the leukocytes and platelets, because these cells may contain higher concentrations of the nutrient than the erythrocytes. After separation, the packed erythrocytes must be washed three times with isotonic saline to remove the trapped plasma, and then homogenized. The latter step is critical because during centrifugation the erythrocytes become density stratified, with younger lighter cells at the top and older denser cells at the bottom.

There is no standard method for expressing the nutrient content of erythrocytes, and each has limitations. The methods used

include nutrient per liter of packed cells, per number of cells, per g of hemoglobin (Hb), or per g of dry material (Vitoux et al., 1999). As an example, erythrocyte folate is expressed as µg/L or nmol/L, whereas erythrocyte zinc is often expressed as µg/g Hb. Concentrations of folate in erythrocytes reflect folate stores(Bailey et al., 2015), whereas results for zinc concentrations in erythrocytes are inconsistent. As a consequence, zinc in erythrocytes is presently not recommended as a biomarker of zinc status by the BOND Expert Panel (King et al., 2015), despite their use in several studies (Lowe et al., 2009).

Erythrocytes can also be used for the assay of a variety of functional biochemical biomarkers based on enzyme systems, especially those depending on B-vitamin-derived cofactors; for more details, see Section 15.4.2. In such cases, the total concentration of vitamin-derived cofactors in the erythrocytes, or the extent of stimulation of specific enzymes by their vitamin-containing coenzymes, is determined. Some of these biomarkers are sensitive to marginal deficiency states and accurately reflect body stores of the vitamin.

15.3.3 Leukocytes

Leukocytes, and some specific cell types such as lymphocytes, monocytes and neutrophils, have been used to monitor medium- to long-term changes in nutritional status because they have a lifespan which is slightly shorter than that of erythrocytes. Therefore, at least in theory, nutrient concentrations in these cell types should reflect the onset of a nutrient deficiency state more quickly than do erythrocytes.

However, several technical factors have limited their use as biomarkers of nutritional status. They include the relatively large volumes of blood required for their analysis, the necessity to process the cells as soon as possible after the specimen is obtained, the difficulties of separating specific leukocytic components from other white blood cell types, and unwanted contaminants in the final cell preparation. Additional technical difficulties may arise if the nutrient content of the cell types varies with the age and size of the cells. In some circumstances, for example during surgery or acute infection, there is a temporary influx of new granulocytes, which alters the normal balance between the cell types in the blood and thus may confound the results. Certain illnesses may also alter the size and protein content of some cell types, and this may also lead to difficulties in the interpretation of their nutrient content(Martin et al., 1993). Hence it is not surprising that results of studies on the usefulness of nutrient concentrations such as zinc in leukocytes or specific cell types as a biomarker of zinc exposure or status have been inconsistent. As a result, zinc concentrations in leukocytes or specific cell types were classified as “not useful” by the Zinc Expert Panel (King et al., 2015).

Detailed protocols for the collection, storage, preparation, and separation of human blood cells are available in Dagur and McCoy (2016). Several methods are used to separate leukocytes from whole blood. They include lysis of erythrocytes, isolating mononuclear cells by density gradient separation, and various non-flow sorting methods. Of the latter, magnetic bead separation can be used to enrich specific cell populations prior to flow cytometric analysis. Lysis of erythrocytes is much quicker than density gradient separation, and results in higher yields of leukocytes with good viability. Nevertheless, density gradient separation methods should be used when purification of cell populations is required rather than simple removal of erythroid contaminants. When flow cytometry is used, cells do not necessarily need to be purified or separated for the study of a particular subpopulation of cells. However, their separation or enrichment prior to flow cytometry does enhance the throughput and ultimately the yield of a desired population of cells.

Again, as noted for erythrocytes, no standard method exists for expressing the content or concentration of nutrients in cells such as leukocytes. Methods that are used include nutrient per unit mass of protein, nutrient concentration per cell, nutrient concentration per dry weight of cells, and nutrient per unit of DNA.

15.3.4 Breast milk

Concentrations of certain nutrients secreted in breast milk — notably vitamins A, D, B_6, B_{12}, thiamin and riboflavin, as well as iodine and selenium — can reflect levels in the maternal diet and body stores (Dror and Allen, 2018). Studies have shown that in regions where deficiencies of vitamin A(Tanumihardjo et al., 2016), vitamin B_{12} (Dror and Allen, 2018), selenium (Valent et al., 2011), and iodine (Dror and Allen, 2018) are endemic, concentrations of these micronutrients in breast milk are low.

In some settings, it is more feasible to collect breast milk samples than blood samples. Nevertheless, sampling, extraction, handling and storage of the breast milk samples must be carried out carefully to obtain accurate information on their nutrient concentrations. To avoid sampling colostrum and transitional milk, which often have very high nutrient concentrations, mature breast milk samples should be taken at least 21d postpartum, when the concentration of most nutrients (except zinc) has stabilized. Ideally, complete 24h breast milk samples from both breasts should be collected, because the concentration of some nutrients

(e.g., retinol) varies during a feed. In community-based studies, however, this is often not feasible. As a result, alternative breast milk sampling protocols have been developed, the choice depending on the study objectives and the nutrient of interest.

To date, only breast milk concentrations of vitamin A have been extensively used to provide information about the vitamin A status of the mother and the breastfed infant (Dror and Allen, 2018a; Dror and Allen, 2018b; Figure 15.2).

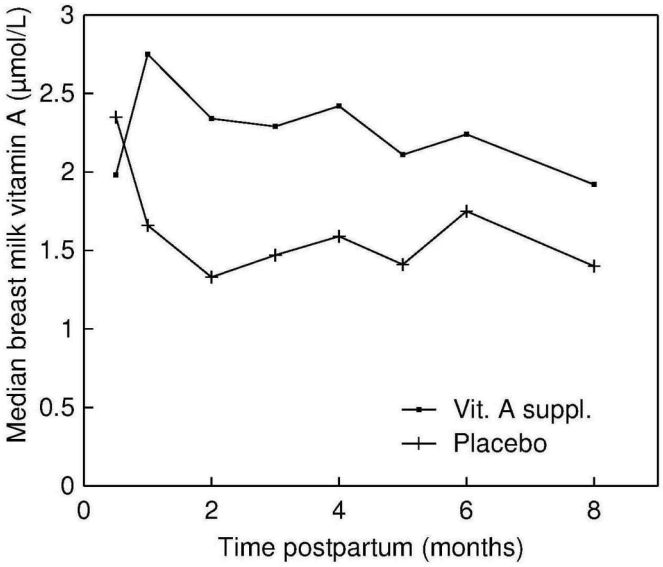


Figure 15.2. Retinol concentrations (μmol/L) in breast milk at baseline and during the subsequent 8mos in supplement and placebo groups. From:

Stoltzfus et al., 1993

, by permission of Oxford University Press.

For the assessment of breast milk vitamin A at the individual level, the recommended practice is to collect the entire milk content of one breast that has not been used to feed an infant for at least 2h, into a dark glass bottle on ice. This procedure is necessary because the fat content of breast milk, and thus the content of fat-soluble vitamin A, increases from the beginning to the end of a single feed (Dror and Allen, 2018). If a full-breast milk sample cannot be obtained, then an aliquot (8–10mL) can be collected before the infant starts suckling, by using either a breast pump or manual self-expression (Rice et al., 2000).

For population-based studies, WHO (1996) suggests collecting random samples of breast milk throughout the day and at varying times following the last feed (i.e., casual samples) in an effort to ensure that the variation in milk fat is randomly sampled. When random sampling is not achievable, the fat-soluble nutrients should be expressed relative to fat concentrations as described in Dror and Allen (2018). The fat content of breast milk can be determined in the field by using the creamatocrit method; details are available in Meier et al. (2006).

Before shipping to the laboratory, the complete breast milk sample from each participant should be warmed to room temperature and homogenized by swirling gently, from which an aliquot of the precise volume needed for analysis can be withdrawn. This aliquot is then frozen at −20°C in an amber or yellow polypropylene tube with an airtight cap, preferably in a freezer without a frost/freeze cycle, until it is analyzed. This strategy of prehomogenization reduces subsequent problems such as attaining uniform mixing after prolonged storage in a freezer.

Table 15.1 compares the performance of breast milk indicators in relation to their ability to detect a response to postpartum vitamin A supplementation in lactating Bangladeshi women (Rice et al., 2000). The most responsive breast milk indicator in this study was the vitamin A content per gram of fat in casual breast milk samples, based on the absolute values of the standardized differences. For more details, see Dror and Allen (2018).

Table 15.1. Response to postpartum vitamin A supplementation measured by maternal and infant indicators. The values shown are m serum retinol data. The means and SDs of the transformed values are presented. [n], number of samples. Data from Rice et al., Ame

Indicator (month post-partum)	Vitamin A group [n]	Plac
Breast milk vit.A (μg/g fat) in casual samples (3 mo)	2.05±0.44 [36]	1.70

Indicator (month post-partum)	Vitamin A group [n]	Plac
Breast milk vit.A (µmol/L) in casual samples (3 mo)	0.12±0.70 [36]	-0.1
Maternal serum retinol (µmol/L) (3 mo)	1.45±0.47 [34]	1.33
Breast milk vit.A (µmol/L) in full samples (3 mo)	-0.33±0.74 [33]	-0.4
Breast milk vit.A (µg/g fat) in full samples (3 mo)	1.87±0.51 [33]	1.82

The analytical methods selected for breast milk should be determined by the chemo-physical properties of the nutrients, their form in breast milk, and their concentrations. Reagents used must be free of adventitious sources of contamination; bound forms of some of the vitamins (e.g., folate, pantothenic acid, vitamins D and B_{12}) must be released prior to extraction and analysis. Increasingly, multi-element mineral analysis is performed by Inductively Coupled Plasma Mass Spectrometry (ICP-MS), whereas for the vitamins, a combination of High-Performance Liquid Chromatography (HPLC) (for thiamin, vitamin A, and vitamin E), ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) (for riboflavin, nicotinamide, pantothenic acid, vitamin B_6, and biotin), and a competitive chemiluminescent enzyme immunoassay (IMMULITE 1000; Siemens) for vitamin B_{12} (cobalamin) are being used (Hampel et al., 2014).

15.3.5 Saliva

Several studies have investigated the use of saliva as a biopsy fluid for the assessment of nutritional status. It is readily available across all ages (newborn to elderly) and collection procedures are noninvasive (unlike blood) so that multiple collections can be performed in the field or in the home.

Steroid and other nonpeptide hormones (e.g., thyroxine, testosterone), some therapeutic and other drugs, and antibodies to various bacterial and viral diseases, can be measured in saliva. The effect of physiological measures of stress such as cortisol and α-amylase on inflammatory biomarkers and immunoglobulin A (IgA) can also be investigated in saliva specimens (Engeland et al., 2019). Studies on the utility of saliva as a biopsy material for metabolomic research are limited. Walsh et al. (2006) reported a high level of both inter- and intra-individual variation in salivary metabolic profiles which was not reduced by standardizing dietary intake on the day before sample collection.

Increasingly, energy expenditure, determined by the doubly labeled water (DLW) method, has been used to assess the validity of reported energy intakes measured using a variety of dietary assessment methods (Burrows et al., 2019). In the DLW method, at least two independent saliva samples, collected at the start and end of the observation interval, are required to measure body water enrichment for ^{18}O and ^{2}H; for more details, see Westerterp (2017).

Some micronutrient concentrations in saliva have also been investigated as a measure of exposure and/or status (e.g., zinc). However, interpreting the results is difficult — results do not relate consistently to zinc intake or status, and suitable certified reference materials and interpretive values for normal individuals are not available. Consequently, the BOND Zinc Expert Panel did not recommend salivary zinc as a biomarker of zinc exposure or status (King et al., 2015).

Saliva is a safer diagnostic specimen than blood; infections from HIV and hepatitis are less of a danger because of the low concentrations of antigens in saliva ((Hofman, 2001). Some saliva specimens, depending on the assay, can be collected and stored at room temperature, and then mailed to the laboratory without refrigeration. However, before collecting saliva samples, several factors must be considered; these are summarized in Box 15.5.

Box 15.5. Factors to be considered when collecting saliva samples

- Is resting or stimulated saliva required? (Stimulated saliva can be collected using sugar-free gum)
- What volume of saliva is required for the assay?
- Is special pretreatment and storage of the saliva required?

- What is the health status of the participants in relation to medications and/or diseases causing a dry mouth?
- Will a quantitative or qualitative assay be performed?

Collection of saliva can be accomplished by expectorating saliva directly into tubes or small paper cups, with or without any additional stimulation. Participants may be requested to rinse their mouth with distilled water prior to the collection. In some cases (e.g., for the DLW method), cotton balls or absorbent pads are used to collect saliva. These can be immersed in a preservative which stabilizes the specimen for several weeks. A disadvantage of this method is that it may contribute interfering substances to the extract and is therefore not suitable for certain analytes.

Alternatively, devices can be placed in the mouth to collect a filtered saliva specimen. These include a small membrane sack that filters out bacteria and enzymes (Saliva Sac; Pacific Biometrics, Seattle, Washington) (Schramm and Smith, 1991), or a tiny plastic tube that contains cyclodextrin to bind the analyte. The latter device, termed the “Oral Diffusion Sink” (ODS), is available from the Saliva Testing and Reference Laboratory, Seattle, Washington (Wade and Haegle, 1991). The ODS device can be suspended in the mouth using dental floss, while the subject is sleeping or performing most of their normal activities with the exception of eating and drinking. In this way, the content of the analyte in the saliva represents an average for the entire collection period.

15.3.6 Sweat

Collection of sweat, like saliva, is noninvasive and can be performed in the field or in the home. Several collection methods for sweat have been used: some are designed to collect whole body sweat, whereas others collect sweat from a specific region of the body, often using some form of enclosing bag or capsule.

Shirreffs and Maughan (1997) have developed a method for collecting whole body sweat involving the person exercising in a plastic-lined enclosure. The method does not interfere with the normal sweating process and overcomes difficulties caused by variations in the composition of sweat from different parts of the body. The method cannot be used for treadmill exercise but can be used for subjects exercising on a cycle ergometer.

A method designed to collect sweat from a specific region of the body involves using a nonocclusive skin patch known as an Osteo-patch. It consists of a transparent, hypo-allergenic, gas-permeable membrane with a cellulose fiber absorbent pad. The patch can be applied to the abdomen or lower back for five days. During the collection period, the nonvolatile components of sweat are deposited on the absorbent pad, whereas the volatile components evaporate through the semipermeable membrane. This method has been used to study collagen cross-link molecules such as deoxypyridinoline in sweat as biomarkers of bone resorption (Sarno et al., 1999).

Potassium levels in sweat are used to normalize the deoxypyridinoline values for variations in sweat volume, as these are highly correlated with sweat output and readily measured by flame atomic emission or ion-selective electrode techniques. Sweat sodium losses can also be measured using an Osteo-patch (Figure 15.3; Dziedzic et al., 2013).

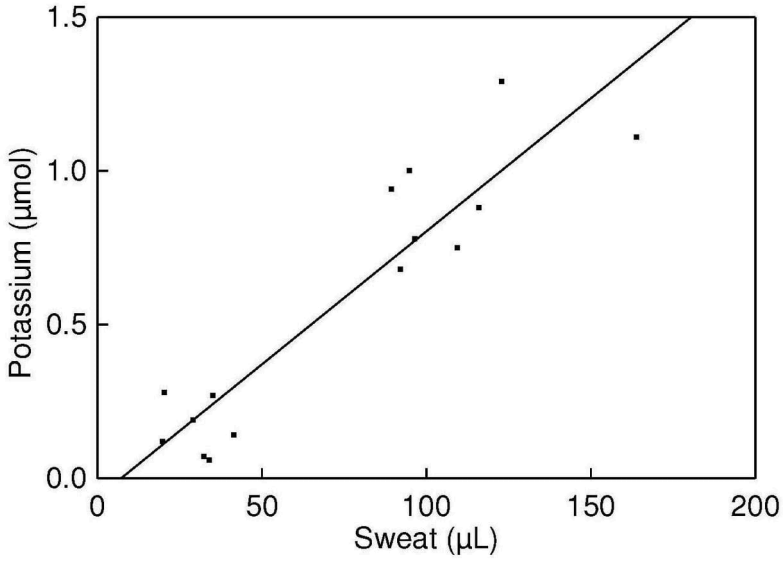


Figure 15.3. Total sweat potassium from patches worn by five healthy subjects vs. total volume of sweat collected. Data from Sarno et al. (1999). Reproduced by permission of OxfordUniversity Press on behalf of the American Society for Nutrition.

A more recent method, known as the Megaduct sweat collector, has been designed for the collection of sweat for mineral analyses (Ely et al., 2012). It appears to avoid skin encapsulation and hidromeiosis (excessive sweating) which may alter sweat mineral concentrations, and captures sweat with mineral concentrations similar to those reported for localized patches.

Differences in the composition of human sweat have been linked, in part, to discrepancies in collection methods. Errors may be caused by contamination, incomplete collection, or real differences induced by the collection procedure.

15.3.7 Adipose tissue

Adipose tissue is a biopsy material that is used in both clinical (Cuerq et al., 2016) and population studies (Dinesen at al., 2018). It can be used as a measure of long-term dietary intake of fat-soluble nutrients, reflecting intakes of certain fatty acids, vitamin E, and carotenoids, all of which accumulate in adipose tissue.

Only fatty acids that are absorbed and stored in adipose tissue without modification, and that are not synthesized endogenously, can be used as biomarkers. Examples of fatty acids that have been used include some specific n-3 and n-6 polyunsaturated fatty acids, trans unsaturated fatty acids, and some odd-numbered and branched-chain saturated fatty acids (e.g., pentadecanoic acid (15:0) and heptadecanoic acid (17:0)). Several other factors that influence the measurement of fatty acid profiles in adipose tissue must also be taken into account; these are summarized in Box 15.6.

Box 15.6. Factors influencing measured fatty acid biomarker levels in adipose tissue. From Arab (2003)

- Dietary intake of the respondent
- Relative amounts of other fatty acids in the adipose tissue samples
- supplement use (such as fish-oil capsules) by the respondent
- Genetic polymorphisms of elongase and desaturase enzymes
- Tissue-sampling site
- Tissue-sampling procedures and subsequent sample handling and storage
- Amount sampled in relation to the analytical method and detection limit
- Lipolysis (the breakdown of fat stored in fat cells)
- Nutritional status (Fe, Zn, Cu and Mg sufficiency)

- Lipogenesis (the production of fat from the metabolism of protein and carbohydrate)
- Diseases: cystic fibrosis, malabsorption, liver cirrhosis, diabetes, Zellweger syndrome
- Fat oxidation abnormalities

The tissue sampling site is also an important consideration when measuring carotenoid concentrations in adipose tissue. Abdominal adipose tissue carotenoid concentrations appear to have the strongest correlation with long-term dietary carotenoid intakes and status (Chung et al., 2009). In contrast, for α-tocopherol, relationships with long-term dietary intakes are independent of adipose tissue site (Schäfer and Overvad, 1990).

Several health outcomes associated with dairy fat consumption have been investigated based on fatty acid concentrations in adipose tissue. As an example, Mozaffarian (2019), in a large pooled analysis of 16 prospective cohort studies in the U.S., Europe, and Australia, showed that higher levels of pentadecanoic acid (15:0), heptadecanoic (17:0), and trans-palmitoleic acid (t16:1n-7) in adipose tissue were associated with a lower risk of type 2 diabetes (Figure 15.4; Imamura et al., 2018).

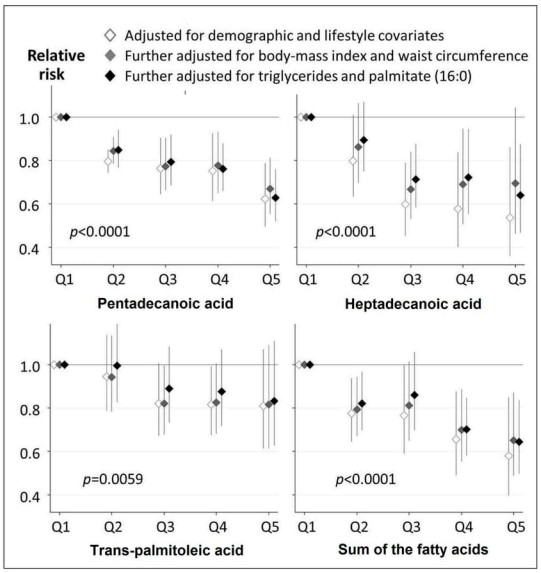


Figure 15.4. Prospective associations of quintile categories of fatty acid biomarkers for dairy fat consumption with the risk of developing type-2 diabetes mellitus. Cohort-specific associations by quintiles were assessed in multivariable models in each cohort and pooled with inverse-variance ± weighted meta-analysis. Cohort-specific multivariable adjustment was made. In the first model (open diamond), estimates were adjusted for sex, age, smoking status, alcohol consumption, socioeconomic status, physical activity, dyslipidaemia, hypertension, and menopausal status for women. Then, the estimates were further adjusted for BMI (grey diamond) and further adjusted for triglycerides and palmitic acid (16:0) as markers of *de novo* lipogenesis (black diamond). To compute p-values for a trend across quintiles, each fatty acid was evaluated as an ordinal variable in the most adjusted model. Redrawn from Imamura et. al, PLoS medicine 15(10), e1002670.

Biomarkers of fatty acids in adipose tissue have also been used to validate the classification of individuals as vegetarian and non-vegetarian in the Adventist Health Study-2, based on the individuals self-reported patterns of consumption of animal and plant-based products (Miles et al., 2019). Results confirmed that the self-reported vegans had a lower proportion of the saturated fatty acids investigated (especially pentadecanoic acid) in adipose tissue, but higher levels of n-6 polyunsaturated fatty acid linoleic (18:2ω-6) and a higher proportion of total ω-3 fatty acids compared to the self-reported non-vegetarians. These trends are consistent with a vegan dietary pattern.

Relationships between long-term dietary intakes of the antioxidant nutrients — α-tocopherol and carotenoids — and their corresponding concentrations in adipose tissue have also been documented in healthy adults. In general, such correlations

exceed those reported between plasma concentrations and diet (Kardinaal et al., 1995; Su et al., 1998). In a large epidemiologic study in which both plasma and adipose tissue carotenoid concentrations were measured, lycopene in adipose tissue(Kohlmeier et al., 1997) but not in plasma (Su et al., 1998) was found to be inversely associated with risk for myocardial infarction.

Simple, rapid sampling methods have been devised for collecting subcutaneous adipose-tissue biopsies, generally from the upper buttock (El-Sohemy et al., 2002), although other sites have also been investigated (Chung et al., 2009). For more discussion on the use of adipose tissue for the assessment of long-term fatty acid and vitamin E status, see Chapters 7 and 18.

15.3.8 Liver and bone

Iron and vitamin A are stored primarily in the body in the liver, and calcium in the bones. Sampling these sites is too invasive for population studies: they are sampled only in research or clinical settings. Dual photon absorptiometry (DXA) is now used to determine total bone mineral content, and is described in detail in Chapter 23.

15.3.9 Hair

Scalp hair has been used as a biopsy material for screening populations at risk for certain trace element deficiencies (e.g., zinc, selenium) and to assess excessive exposure to heavy metals (e.g., lead, mercury, arsenic). Detailed reviews are available from the IAEA (1993; 1994). Caution must be used when interpreting results for hair mineral analysis from commercial laboratories because results can be unreliable (Hambidge, 1982; Seidel et al., 2001; Mikulewicz et al., 2013).

Hair incorporates trace elements and heavy metals into the matrix when exposed to the blood supply during synthesis within the dermal papilla. When the growing hair approaches the skin surface, it undergoes keratinization and the trace elements accumulated during its formation become sealed into the keratin protein structures and isolated from metabolic processes. Hence, the trace element content of the hair shaft reflects the quantity of the trace elements available in the blood supply at the time of its synthesis, not at the time of sampling (Kempson et al., 2007).

Analysis of trace element levels in hair has several advantages compared to that of blood or urine; these are summarized in Box 15.7.

Box 15.7. Some of the advantages of hair as a biopsy material

- Higher concentrations of trace elements are found in hair, relative to blood or urine, making analysis easier; results for the ultra-trace elements such as chromium and manganese are more consistent.
- Concentrations are more stable and hair trace element levels are not subject to the rapid fluctuations associated with diet, diurnal variation, and so on.
- No trauma is involved in the collection of hair samples.
- No special preservatives are needed, and samples can be stored in plastic bags at room temperature without deterioration.

Nevertheless, a major limitation of the use of scalp hair is its susceptibility to exogenous contamination. Hopps (1977) noted that sweat from the eccrine sweat glands may contaminate the hair with elements derived from body tissues. Other exogenous materials that may modify the trace element composition of hair include air, water, soap, shampoo, lacquers, dyes, and medications. Selenium in antidandruff shampoos, for example, significantly increases hair selenium content, and the selenium cannot be removed by standardized hair-washing procedures (Davies, 1982). For other trace elements, results from hair-washing procedures have been equivocal. Some (Hilderbrand and White, 1974), but not all (Gibson and Gibson, 1984), investigators have observed marked changes in hair trace element concentrations after hair cosmetic treatments. The relative importance of these sources remains uncertain, and standardized procedures for hair sampling and washing prior to analysis are essential.

The currently recommended hair sampling method is to use the proximal 10–20mm of hair, cut at skin level from the occipital portion of the scalp (i.e., across the back of the head in a line between the top of the ears) with stainless steel scissors. This procedure, involving the sampling of recently grown hair, minimizes the effects of abrasion of the hair shaft and exogenous contamination. In addition, the specimens collected in this way will reflect the uptake of trace elements or heavy metals by the follicles 4–8 weeks prior to sample collection provided that the rate of hair growth has been normal. Before washing the hair specimens to remove exogenous contaminants such as atmospheric pollutants, water and sweat, any nits and lice should be

removed under a microscope or magnifying glass where necessary, using Teflon-coated tweezers. For each sample details of the ethnicity, age, sex, hair-color, height, weight, season of collection, smoking, presence of disease states including malnutrition, and use of antidandruff shampoos or cosmetic treatments, should always be recorded to aid in the interpretation of the data.

Some investigators suggest that the rate of hair growth influences hair trace element concentrations. Scalp hair grows at about 1cm/mo, but in some cases of severe protein-energy malnutrition (Erten et al., 1978) and the zinc deficiency state acrodermatitis enteropathica (Hambidge et al., 1977), growth of the hair is impaired. In such cases, hair zinc concentrations may be normal or even high. No significant differences, however, were observed in the trace element concentrations of scalp and pubic hair samples (DeAntonio et al., 1982), despite marked differences in the rate of hair growth at the two anatomical sites. These results suggest that the relative rate of hair growth is not a significant factor in controlling hair trace element levels.

Several different washing procedures have been investigated, including the use of nonionic or ionic detergents, followed by rinsing in distilled or deionized water to remove absorbed detergent. Various organic solvents such as hexane-methanol, acetone, and ether, have also been recommended, either alone or in combination with a detergent (Salmela et al., 1981). Washing with nonionic detergents (e.g., Triton X-100) (with or without acetone) is preferred as nonionic detergents are less likely to leach bound trace minerals from the hair and yet are effective in removing superficial adsorbed trace elements. Washing with chelating agents such as EDTA should be avoided because of the risk of removing endogenous trace minerals from the hair shaft (Shapcott, 1978).

After washing and rinsing, the hair samples must be vacuum- or oven-dried depending on the chosen analytical method, and stored in a desiccator prior to laboratory analysis. When the traditional analytical methods such as flame Atomic Absorption Spectrophotometry (AAS) or multi-element Inductively- Coupled Plasma Mass Spectrometry (ICP-MS) are used, washed hair specimens must be prepared for analysis using microwave digestion, or wet or dry ashing. In the future, tetramethylammonium hydroxide (TMAH) to solubilize hair at room temperature may be used, eliminating time-consuming ashing or wet digestion (Batista et al., 2018). Non-destructive instrumental neutron activation analysis (INAA) can also be used, when the washed hair specimens are placed in small, weighed, TE-free, polyethylene bags or tubes, and oven dried for 24h at 55°C. After cooling in a desiccator, the packaged specimens are sealed and weighed, prior to irradiation in a nuclear reactor.

A Certified Reference Material (CRM) for human hair is available (e.g., Community Bureau of Reference, Certified Reference Material no. 397) from the Institute for Reference Materials and Measurements, Retieseweg, B-2440 Geel, Belgium. Currently, interpretation of hair trace element concentrations for screening populations at risk of deficiency is limited by the absence of universally accepted reference values. For a detailed step-by-step guide to measuring hair zinc concentrations, the reader is advised to consult IZiNCG (2018).

In summary, more data on other tissues from the same individuals are urgently required to interpret the significance of hair trace element concentrations. Hair is certainly a very useful indicator of the body burden of heavy metals such as lead, mercury, cadmium and arsenic. It is also valuable in the case of selenium and chromium, and possibly zinc. Data for other elements such as iron, calcium, magnesium, and copper should be interpreted with caution (Seidel et al., 2001).

15.3.10 Fingernails and toenails

Nails have been investigated as biopsy materials for trace element analysis (Bank et al., 1981; van Noord et al., 1987). Nails, like hair, also incorporate trace elements into the nail matrix when it is exposed to the blood supply within the nail matrix germinal layer, and thus reflect the quantity of trace elements available in the blood supply at the time of nail synthesis (He, 2011). During the growth of the nail, the proliferating cells in the nail germinal layer are converted into horny lamellae. Nails grow more slowly than hair at rates ranging from 1.6mm/month for toenails to 3.5mm/month for fingernails, and, like hair, are easy to sample and store. In cases where nail growth is arrested, as may occur in onychophagia (compulsive nail biting), nails should not be used (He, 2011).

The elemental composition of toenails has been used as a long-term biomarker of nutritional status for some elements, notably selenium. Selenium concentrations in toenails correlate with geographic differences in selenium exposure (Figure 15.5), (Morris et al., 1983; Hunter et al., 1990).

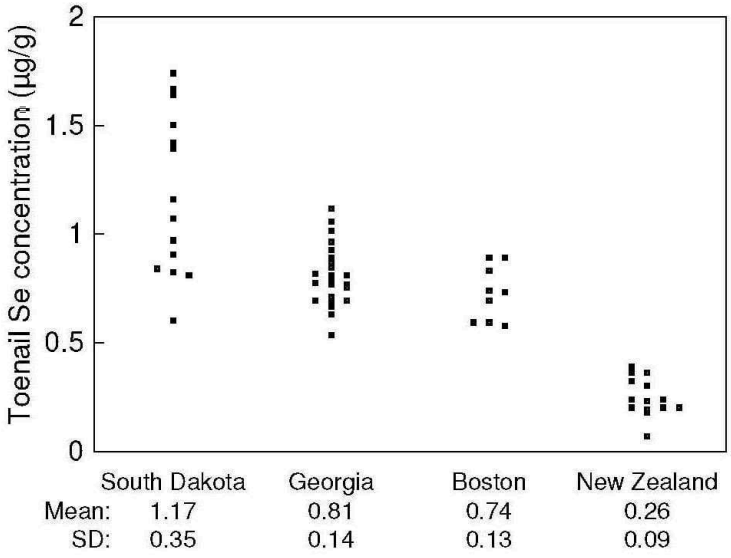


Figure 15.5. Toenail selenium concentrations in a high-selenium area (South Dakota), Georgia, and Boston, compared with a low-selenium area (New Zealand). Redrawn fromMorris et al., 1983).

At the individual level, concentrations of selenium in toenails correlate with those in habitual diets, serum, and whole blood (Swanson et al., 1990).

In a recent study of young children in Laos, nail zinc concentrations were higher at endline in those children receiving a daily preventive zinc supplement (7–10mg Zn/d) for 32–40 weeks compared to those given a therapeutic zinc dose (20g) for only 10d (geometric mean, 95% CI) (115.8, 111.6–119.9 vs. 110.4, 106.0–114.8µg/g; p=0.055) (Wessells et al., 2020). Nail zinc concentrations have also been used as a longer-term retrospective measure of zinc exposure in case-control studies. For example, in a prospective study of U.S. urban adults (n=3,960), toenail zinc was assessed in relation to the incidence of diabetes, although no significant longitudinal association was found (Park et al., 2016).

The elemental composition of nails is influenced by age, possibly sex, rate of growth, onychophagia (compulsive nail biting), geographical location, and possibly by disease states (e.g., cystic fibrosis, Wilson's disease, Alzheimer's disease, and arthritis) (Takagi et al., 1988; Vance et al., 1988). Environmental contamination and chemicals introduced by nail polish could be a potential problem, unless they are removed by washing (He, 2011). Bank et al. (1981) recommend cleaning fingernails with a scrubbing brush and a mild detergent, followed by mechanical scraping to remove any remaining soft tissue before clipping. Nail samples should then be washed in aqueous non-ionic detergents rather than organic solvents, and dried under vacuum prior to preparation and analysis by the same traditional analytical techniques as are used for hair specimens. Tetramethylammonium hydroxide (TMAH) can also be used to solubilize nails at room temperature, eliminating time-consuming ashing or wet digestion, thus enhancing sample throughput (Batista et al., 2018).

For non-destructive analytical methods such as instrumental neutron activation analysis (INAA) and the newer technique involving laser-induced breakdown spectroscopy (LIBS), cleaning fingernail clippings with acetone (analytical grade) in an ultrasonic bath for 10min followed by drying in air for20–30min is recommended (Riberdy et al., 2017). Preliminary results suggest that *in situ* measurement of fingernail zinc by LIBS has potential as a non-invasive, convenient screening tool for identifying zinc deficiency in populations, but may lack the precision required to generate absolute concentrations for individuals (Riberdy et al., 2017). A non-destructive portable X-ray fluorescence system has also been used to explore the measurement of zinc in a single nail clipping; more studies are needed to establish its usefulness (Fleming et al., 2020).

Unlike hair, no Standard Reference Materials presently exist for nail trace element analysis. Instead, in-house controls prepared from homogenous pooled samples of powdered fingernails and toenails can be prepared and spiked with several different known quantities of the trace element of interest and the recoveries measured. Alternatively, an aliquot of the in-house control can be sent to a reputable laboratory and the results compared. Likewise, there are no universally accepted reference values for nail trace element concentrations, limiting their use for assessing risk of trace element deficiencies in populations. More studies comparing the trace element composition of fingernails and toenails with corresponding concentrations in other biomarkers of body tissues and fluids, as well as habitual dietary intakes, are needed before any definite recommendations on the use of fingernails or

toenails as a biomarker of exposure or status can be made.

15.3.11 Buccal mucosal cells

Buccal mucosal cells have been investigated as a biopsy sample for assessing α-tocopherol status (Kaempf et al., 1994; Chapter 18) and dietary lipid status (McMurchie et al., 1984; Chapter 7), but interpretive criteria to assess these results are not available. These cells have also been explored as a biomarker of folate status (Johnson et al., 1997), although smoking is a major confounder as a localized folate deficiency is generated in tissues exposed to cigarette smoke (Piyathilake et al., 1992). Buccal mucosal cells are also increasingly used in epidemiological studies that involve DNA (Potischman, 2003).

Buccal mucosal cells can be sampled easily and noninvasively by gentle scraping with a spatula. Cells must be washed with isotonic saline prior to sonication and analysis. Contamination of buccal cells with food is a major problem, however, and has prompted research into new methods for the collection of buccal mucosal cells.

15.3.12 Urine

If renal function is normal, biomarkers based on urine or the urinary excretion rate of a nutrient or its metabolite can be used to assess exposure or status for some trace elements (e.g., chromium, iodine, selenium), the water-soluble B-complex vitamins, and vitamin C. The method depends on the existence of a renal conservation mechanism that reduces the urinary excretion of the nutrient or metabolite when body stores are depleted. Urine cannot be used to assess the status of the fat-soluble vitamins A, D, E, and K, as metabolites are not excreted in proportion to the amount of these vitamins consumed, absorbed, and metabolized.

Urinary excretion can also be used to measure exposure to certain nutrients, as well as some food components and food groups. Isaksson (1980) was one of the first investigators to use urinary nitrogen excretion levels in single 24h urine samples to estimate exposure to protein intakes from a 24h food record. Since that time, several urinary biomarkers for other nutrients, and for certain food components and food groups, have been investigated, in some cases as biomarkers of exposure or status, as noted in Section 15.2.

Urinary excretion assessment methods almost always reflect recent dietary intake or acute status, rather than chronic nutritional status. If information on long-term exposure is required, multiple 24h urine samples collected over a period of weeks should be used. For example, to obtain a stable measurement of long-term exposure to sodium, potassium, calcium, phosphate and magnesium, three 24h urine samples from healthy adults spaced over a predefined time period are required (Sun et al., 2017).

For some of the water-soluble vitamins (e.g., thiamin, riboflavin and vitamin C), the amount excreted depends on both the nutrient saturation of tissues and on the dietary intake. Furthermore, urinary excretion tends to reflect intake when intakes of the vitamins are moderate to high relative to the requirements, but less so when intakes are habitually low. In other circumstances such as infections, trauma, the use of antibiotics or medications, and conditions that produce negative balance, increases in urinary excretion may occur despite depletion of body nutrient stores. For example, drugs with chelating abilities, alcoholism, and liver disease can increase urinary zinc excretion, even in the presence of zinc deficiency.

For measurement of a nutrient or a corresponding metabolite in urine, it is essential to collect a clean, properly preserved urine sample, preferably over a complete 24h period. Thymol crystals dissolved in isopropanol are often used as a preservative (Mente et al., 2009). For nutrients that are unstable in urine (e.g., vitamin C), acidification and cold storage are required to prevent degradation.

To monitor the completeness of any 24h urine collection, urinary creatinine excretion is often measured (Chapter 16). This approach assumes that daily urinary creatinine excretion is constant for a given individual, the amount being related to muscle mass. In fact, this excretion can be highly variable within an individual (Webster and Garrow, 1985), and varies with age (Yuno et al., 2011).

Box 15.8. Possible reasons (other than the under-collection of 24h urine samples) for low PABA recovery values

- Failure to take all three PABA tablets

- Taking tablets late in the evening with a large meal that reduces gastric emptying time and uptake in the intestine
- Impaired renal function
- Errors in preparation of urine aliquots
- Analytical errors

Estimates of the within-subject coefficient of variation for creatinine excretion in sequential daily urine collections range from 1% to 36% (Jackson, 1966; Webster and Garrow, 1985). Hence, creatinine determinations may detect only gross errors in 24h urine collections (Bingham and Cummings, 1985).

British investigators have used an alternative marker, paraaminobenzoic acid (PABA), to assess the completeness of urine collections (Bingham and Cummings, 1985). Para-aminobenzoic acid is taken in tablet form with meals — one tablet of 80mg PABA three times per day. It is harmless, easy to measure, and rapidly and completely excreted in urine.

Possible explanations for low PABA recovery values besides the under-collection of urine samples are summarized in Box 15.8. Studies have shown that any urine collection containing less than 85% of the administered dose is probably incomplete (Bingham and Cummings, 1985), suggesting that PABA is a useful marker for monitoring the completeness of urine collection. The incomplete nature of urine collections with a mean PABA recovery of < 79% is emphasized in Figure 15.6.

Bingham, S., & Cummings, J. H. (1983). The use of 4-aminobenzoic acid as a marker to validate the completeness of 24 h urine collections in man. Clinical science (London, England : 1979), 64(6), 629–635.

A method has been devised for adjusting urinary concentrations of nitrogen, sodium and potassium in cases where the recovery of PABA is between 50% and 80%. It is based on the linear relationship between the PABA recovery and the amount of analytes in the urine, as shown in Figure 15.7, and allows the use of incomplete 24h urine collections. However, this adjustment method is not recommended in cases where PABA recovery is below 50% Figure 15.7.

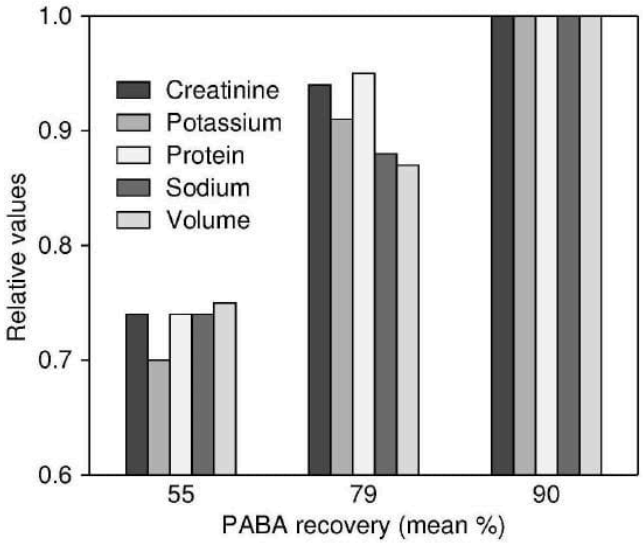


Figure 15.6. The relationship of urinary PABA recovery and urinary creatinine, potassium, protein (derived from nitrogen), sodium, and volume in three groups of patients with median PABA recovery of 55% (n=28), 79% (n=24), and 90% (n=21). The urinary variables are expressed in relative terms in relation to the highest PABA recovery group, which is set to 1.0. Data from Johansson et al., 1998. reproduced with permission of Cambridge University Press.

Several investigators have measured urinary biomarker concentrations of nitrogen, sodium and potassium to validate dietary intakes in population studies, some of which assessed the completeness of 24h urine collection by analysis of PABA concentration in the urine. For example, Wark et al. (2018) assessed the validity of intakes in adults (n=212) of protein, sodium and potassium estimated from 3 × 24h recalls taken 2 weeks apart using an online 24h recall tool (myfood24) by comparison with urinary biomarkers.

Participants were instructed to take one 80mg PABA tablet with each of three meals during the 24h urine collection period, and urinary concentrations for nitrogen, sodium and potassium were then adjusted for completeness of urine samples when PABA recovery was 50–85%. The investigators calculated that 93% of PABA, 81% of nitrogen, 86% of sodium and 80% of potassium were excreted within 24h. Table 15.2. shows the geometric mean and 95% confidence interval (CI) for protein, potassium and sodium intake and associated nutrient densities as assessed by myfood24 online recall and the biomarkers relating to the first clinic visit. Estimates of intake from myfood24 were similar to the biomarker measurements for protein, but higher for both potassium and sodium. Such discrepancies may be attributed to reporting error, daily variation in diet, and limitation of food composition tables, especially for sodium as a result of addition of discretionary salt to foods in manufacture or discretionary salt at the table.

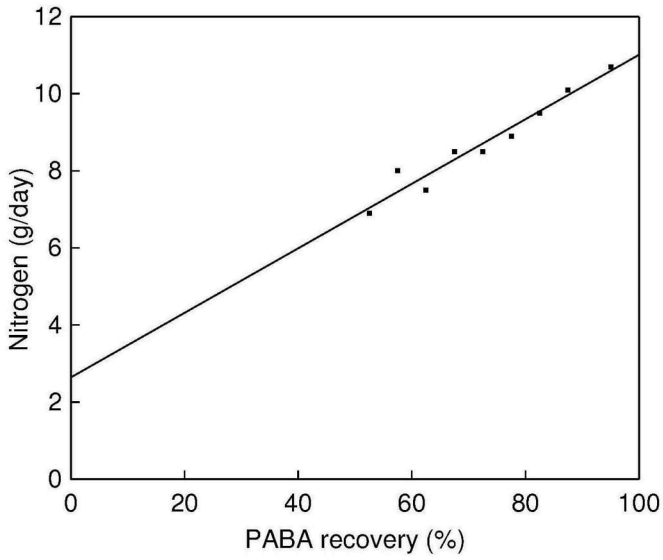


Figure 15.7. The relationship between PABA recovery (%) and the nitrogen output in urine (g/d). The PABA recovery values have been classified into 5% intervals from 50% to 90% and one interval between 90% and 110%. The number of subjects in each interval is 10 or more. Total n=312, r^2=0.9752. Data from: Johansson et al., 1999,

Table 15.2. Geometric means and 95% confidence interval (CI) for protein, potassium, sodium and total sugar intake and density as assessed by myfood24 and biomarkers relating to the first clinic visit. Nutrient density for protein, potassium, sodium and total sugars is expressed in g/MJ of total energy intake. n is the number of participants who had both the dietary assessment measure and the biomarker. Data from Wark et al., BMC medicine, 16(1), 136.

	myfood24		Biomarker/reference tool	
	n	Geometric mean (95% CI)	n	Geometric mean (95% CI)
Nutrient intake:				
Protein (g)	208	70.5 (66.1, 75.2)	192	68.4 (64.1, 72.8)
Potassium (g)	208	2.7 (2.5, 2.9)	192	2.1 (1.9, 2.3)
Sodium (g)	208	2.3 (2.1, 2.5)	192	1.8 (1.7, 2.0)
Nutrient density:				
Protein (g/MJ)	208	9.5 (9.0, 9.9)	180	6.2 (5.8, 6.7)
Potassium (g/MJ	208	0.36 (0.35, 0.38)	180	0.19 (0.18, 0.21)
Sodium (g/MJ)	208	0.31 (0.29, 0.33)	180	0.16 (0.15, 0.18)

Twenty-four-hour urine samples can be difficult to collect in non-institutionalized population groups. Instead, first-voided fasting morning urine specimens are often used, as they are less affected by recent dietary intake. Such specimens were used in the U.K.

National Diet and Nutrition Survey of young people 4–18y (Gregory et al., 2000). Special Bori-Vial vials containing a small amount of boric acid as a preservative can be used for the collection of first-voided fasting samples. Sometimes, only nonfasting casual urine samples can be collected. Such casual urine samples are not recommended for studies at the individual level, because concentrations of nutrients and metabolites in such samples are affected by liquid consumption, recent dietary intake, body weight, physical activity and other factors.

When first-voided fasting or casual urine specimens are collected, urinary excretion is sometimes expressed as a ratio of the nutrient to urinary creatinine in an effort to correct for both diurnal variation and fluctuations in urine volume. For some urinary biomarkers, specific gravity has been used to correct for urine volume in casual urine samples rather than urinary creatinine (Newman et al., 2000).

As a biomarker of recent exposure to iodine at the population level, WHO / UNICEF / ICCIDD, 2007 recommend collecting casual urine samples and expressing the results in terms of the population median urinary iodine concentration (µg/L). A median urinary iodine concentration of 100–199µg/L in school-age children, for example, indicates adequate iodine nutrition. However, this does not quantify the percentage of individuals with habitually deficient or excessive intakes of iodine.

Daily iodine intake can be calculated from urinary iodine based on the following assumptions: over 90% of iodine is excreted in the urine in the subsequent 24–48h; median 24h urine volume is about 0.0009L/h/kg; average bioavailability of iodine in the diets is 92%. Therefore:

$$\begin{aligned} \text{Iodine intake} &= 0.0009 \times 24/0.92 \times \text{Wt} \times \text{Ui} \\ &= 0.0235 \times \text{Wt} \times \text{Ui} \end{aligned}$$

where Wt is the body weight (kg) and Ui is the urinary iodine (µg/L).

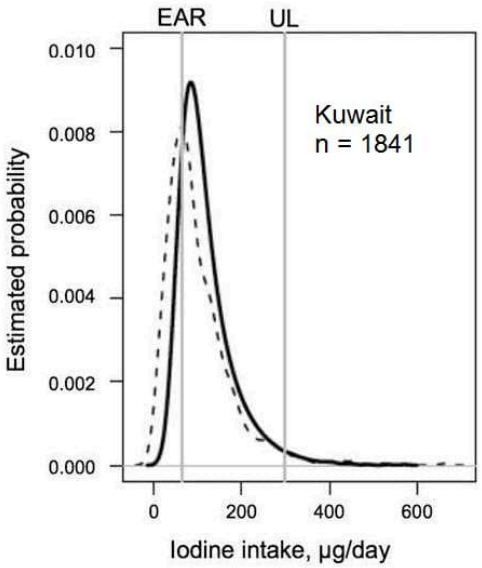


Figure 15.8. The distribution of iodine intakes among children aged 4–8y derived from a single spot urine sample (broken line) and after adjustment for within- and between-subject variation (unbroken line). The Estimated Average Requirement (EAR) for children aged 4–8y is 65 and the Tolerable Upper Intake Level (UL) is 300µg/d. Data from Zimmermann et al., 2016, by permission of Oxford University Press.

This equation has been applied to calculate daily iodine intakes of children based on casual urinary iodine concentrations collected during national surveys in Kuwait, Oman, Thailand, and Qatar, and during a regional study in China. In these surveys, a second repeat casual urine sample was collected in a random subsample of the children on a nonconsecutive day (Figure 15.8). This permits an adjustment to be made to the observed distribution of iodine intakes to remove the variability introduced by day-to-day variation in iodine intakes within an individual (i.e., to remove the within-subject variation) using specialized software, in this case the Iowa State University method (Carriquiry, 1999). For more details of this adjustment method, see Chapter 3.

Figure 15.8 shows that the adjustment process yields a distribution with reduced variability that preserves the shape of the original observed distribution. The adjusted distribution can then be used to predict the proportion of the population at risk of inadequate or excessive intakes of iodine using the Estimated Average Requirement (EAR) / Tolerable Upper Level (UL) cutoff point method; see Chapter 8b for more details. Note that the proportion of children classified with inadequate intakes in both

Kuwait and China was markedly lower based on the adjusted distribution of intakes compared to the unadjusted distribution (Table 15.3).

Table 15.3. Prevalence of inadequate iodine intake by the EAR and UL cutoff method with the use of internal ("true") variance estima Values are means ± SEs. Age groups of children correspond to the U.S. DRI groups.

Age group of children	Unadjusted prevalence below the EAR	True prevalence below the EAR, adjusted wit internal variance
4–8y		
Kuwait	35.3 ± 1.7	19.4 ± 5.7
Oman	24.3 ± 1.8	7.5 ± 4.7
China	20.5 ± 2.5	10.1 ± 4.4
9–13y		
Kuwait	30.9 ± 1.4	17.4 ± 3.6
Oman	18.6 ± 1.1	10.5 ± 2.1
China	24.0 ± 3.9	3.5 ± 7.3

This page titled 13.3: Biomarkers of status (15.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13.4: Biomarkers of function (15.4)

Functional biomarkers can be subdivided into two groups: functional biochemical, and functional physiological or behavioral, biomarkers. They measure the extent of the functional consequences of a specific nutrient deficiency and hence have greater biological significance than the static biomarkers, as noted earlier some functional biomarkers are also being used as substitutes for chronic disease outcomes, when they are termed “surrogate biomarkers” (Yetley et al., 2017).

Functional biochemical biomarkers serve as early biomarkers of subclinical deficiencies. They may involve the measurement of an abnormal metabolic product in blood or urine samples arising from a deficiency of a nutrient- dependent enzyme. Alternatively, for some nutrients, reduction in the activity of enzymes that require a nutrient as a coenzyme or prosthetic group can be measured. For example, the activity of erythrocyte glutamic oxaloacetic transaminase has been reported to better reflect the intake of vitamin B_6 than the plasma concentrations of pyridoxal phosphate, especially in adults < 65y (Elmadfa and Meyer, 2014).

Changes in blood components related to intake of a nutrient can also be determined, and load or tolerance tests conducted on individuals *in vivo*. Sometimes, tissues or cells are isolated and maintained under physiological conditions for biomarkers of *in vivo* functions. Biomarkers related to host defense and immunocompetence are the most widely used of this type. For some of the nutrients (e.g., niacin), functional biochemical biomarkers may not be available.

In research settings, stable isotope techniques are used to measure the size of the body pool(s) of a nutrient (e.g., the vitamin A content of the liver; see Chapter 18), and for kinetic modeling to assess the integrated whole-body response to changes in nutrient status (e.g., protein, copper, zinc). The latter approach is especially useful for detecting subtle changes that may not be responsive to static indices (King et al., 2000). Figure 15.9 shows a marked reduction in the endogenous fecal excretion of zinc over a 6mo period on a low-zinc diet. Such a reduction can be quantified only with isotopic techniques.

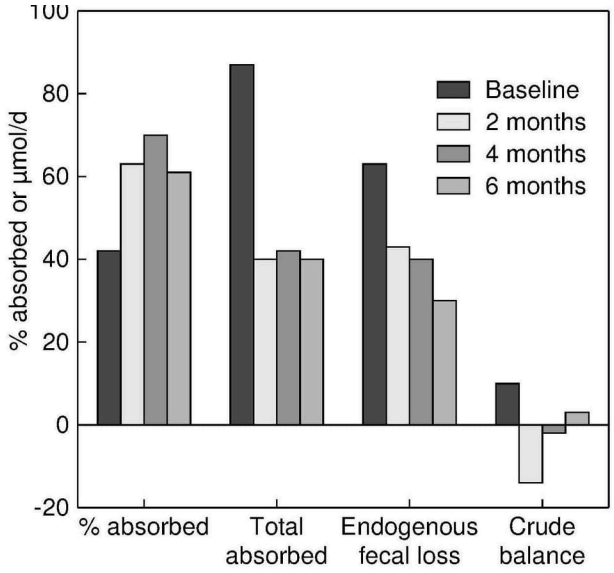


Figure 15.9. Changes in intestinal zinc absorption and endogenous loss over 6mos while on a low-zinc intake of 63µmol/d (4.1mg/d) Data from King et al., 2000, by permission of Oxford University Press.

New molecular techniques are now used in research to measure, for example, mRNA for proteins (e.g., metallothionein), the expression of which is regulated by metal ions such as zinc (Hirschi et al., 2001). Correlations between biomarkers of DNA damage and micronutrient status are also being investigated in view of the growing knowledge of their roles as cofactors or as components of DNA repair enzymes. For example, marginal zinc depletion impairs DNA repair and increases the number of DNA strand breaks. However, these breaks are not specific markers for zinc depletion as insufficient intakes of choline, folate, and niacin also cause an increase in DNA strand breaks (Zyba et al., 2017). Genetic variation can now be identified through DNA testing, and when used in combination with nutritional biomarkers, can assist in understanding variations in metabolism and in identifying subpopulations at risk of disease; see Section 15.7.

Most functional physiological and behavioral biomarkers are less invasive, often easier to perform, and more directly related to disease mechanisms or health status than are functional biochemical biomarkers. In general, however, functional physiological or behavioral biomarkers are not very sensitive or specific and must be interpreted in conjunction with more specific nutrient biomarkers. As noted earlier, these functional physiological and behavioral biomarkers often measure the net effects of contextual factors that often include social and environmental factors as well as nutrition.

Disturbances in these biomarkers are generally associated with more prolonged and severe nutrient deficiency states or in some circumstances, risk of chronic diseases (Yetley et al., 2017). Examples include measurements of impairments in growth, of response to vaccination (as a biomarker of immune function), and of vision, motor development, cognition, depression and high blood pressure, all of which are less invasive and easier to perform. Some important examples of functional biomarkers include the following:

Functional biochemical biomarkers

- Abnormal concentrations of metabolic products in blood or urine arising from reduced activity of a nutrient-dependent enzyme (e.g., urinary excretion of xanthurenic acid, formiminoglutamic acid (FIGLU), and methylmalonic acid as a test of vitamin B_6 and vitamin B_{12} deficiency.
- Changes in enzyme activities that depend on a given nutrient (e.g., erythrocyte glutathione reductase activity for riboflavin; erythrocyte transketolase activity for thiamin; erythrocyte glutamic oxaloacetic transaminase for vitamin B_6).
- Changes in blood components (e.g., whole blood hemoglobin for iron assessment; thyroglobulin for iodine status; retinol-binding protein for vitamin A; holotranscobalamin for vitamin B_{12}).

Functional physiological and behavioral biomarkers

- *In vitro* tests of *in vivo* functions (e.g., lymphocyte proliferation for protein-energy, zinc, and iron).
- Load and tolerance tests and induced responses *in vivo* (e.g., Relative Dose Response: load test for vitamin A; CobaSorb test: load test to assess vitamin B_{12} absorption).
- Induced responses *in vivo* (e.g., delayed-type hypersensitivity, often used to identify protein-energy malnutrition).
- Spontaneous *in vivo* responses (e.g., dark adaptation / vision at low intensity for vitamin A; taste acuity for zinc; handgrip strength for lower-body strength).
- Growth or developmental responses (e.g., growth velocity for protein-energy, zinc, etc.; cognitive performance for iron, iodine, vitamin D, folate, and vitamin B_{12}; motor development for micronutrients; depression for folate and zinc).

15.4.1 Abnormal metabolic products in blood or urine

Many of the vitamins and minerals act as coenzymes or as prosthetic groups for enzyme systems. During deficiency, the activities of these enzymes may be reduced, resulting in the accumulation of abnormal metabolic products in the blood or urine.

Xanthurenic acid excretion in urine, together with other tryptophan metabolites, is elevated in vitamin B_6 deficiency because the activity of kynureninase in the tryptophan-niacin pathway is reduced. This leads to the increased formation and excretion in the urine of xanthurenic acid and other tryptophan metabolites, including both kynurenic acid and 3-hydroxyl-kynurenine. Determination of urinary xanthurenic acid is usual because it is easily measured.

Plasma homocysteine concentrations are elevated in both vitamin B_{12} and folate deficiency. In vitamin B_{12} deficiency, when levels fall below 300pmol/L, the activity of methionine synthase, an enzyme that requires vitamin B_{12}, is reduced. This enzyme catalyzes the remethylation of homocysteine to methionine. Hence, reduction in the activity of methionine synthase leads to increases in plasma homocysteine concentrations (Allen et al., 2018). The remethylation pathway of homocysteine to methionine is also dependent on folate, so when folate status is low or deficient, then plasma homocysteine is generally elevated (Bailey et al., 2015). Therefore, in folate or vitamin B_{12} deficiency, homocysteine accumulates and concentrations in plasma increase. Measurement of plasma homocysteine as a sensitive functional biomarker of low folate status has been recommended by the BOND Folate Expert Panel. However, they highlight its poor specificity because it is elevated with other B-vitamin deficiencies besides folate and vitamin B_{12}(including vitamin B_6, and riboflavin), with lifestyle factors, with renal insufficiency, and with drug treatments (Bailey et al., 2015).

Elevated circulating homocysteine concentrations have been associated with an increased risk of hypertension, cardiovascular

disease, and cerebrovascular disease based on observational studies. Several mechanisms have been proposed whereby hyperhomocysteinemia may mediate risk of these diseases. Details of the collection and analyses of plasma samples for homocysteine are available in Bailey et al. (2015).

Methyl malonic acid (MMA) concentrations in plasma or urine are elevated in vitamin B_{12} deficiency but unaffected by folate or other B vitamins. Vitamin B_{12} serves as a cofactor for the enzyme methylmalonic-CoA mutase. This enzyme is required for the conversion of methylmalonyl-CoA to succinyl-CoA. Methylmalonic acid (MMA) is a side reaction product of methylmalonyl-CoA metabolism, and increases with vitamin B_{12} depletion. Concentrations of MMA reflect B_{12} stores rather than recent B_{12} intake and are considered a relatively specific and sensitive biomarker of vitamin B_{12}status by the BOND Vitamin B_{12} Expert Panel (Allen et al., 2018).

In serum or urine, MMA concentrations reflect the adequacy of B_{12} status for the biochemical function of the enzyme methylmalonic-CoA mutase, which is required for the conversion of methylmalonyl-CoA to succinyl-CoA. MMA, usually a side reaction product of methylmalonyl CoA metabolism, increases with B_{12} depletion (Allen et al., 2018). If urinary MMA is to be measured and the collection of 24h urine samples is not feasible, then urinary creatinine should also be assayed to correct for variability in urine concentration and results expressed as per mg or mmol creatinine.

For more information on elevated levels of homocysteine and MMA, readers are advised to consult the two BOND reports: Bailey et al. (2015) and Allen et al. (2018).

15.4.2 Reduction in activity of enzymes

Methods that involve measuring a change in the activity of enzymes which require a specific nutrient as a coenzyme or prosthetic group are generally the most sensitive and specific. Often the enzyme is associated with a specific metabolic defect and associated nutrient deficiency (e.g., lysyl oxidase for copper, aspartate aminotransferase for vitamin B_6, glutathione reductase for riboflavin, transketolase for thiamin).

The activity of the enzyme is sometimes measured both with and without the addition of saturating amounts of the coenzyme added *in vitro*. The *in vitro* stimulation of the enzyme by the coenzyme indicates the degree of unsaturation of the enzyme, and therefore a measure of deficiency. When nutritional status is adequate, the added coenzyme has little effect on the overall enzyme activity, so the ratio of the two measurements is very close to unity. However, when a deficiency exists, the added coenzyme increases enzyme activity to a variable extent, depending on the degree of deficiency. Such tests, often termed "enzyme stimulation tests", may be used for vitamin B_6, riboflavin and thiamin, and employ the activities of aminotransferases, glutathione reductase and transketolase respectively. Erythrocytes are used for these enzyme stimulation tests because erythrocytes are particularly sensitive to marginal deficiencies and provide an accurate reflection of body stores for vitamin B_6, riboflavin, and thiamin. Indeed, such vitamin-deficient erythrocytes may respond to supplements of B_6, riboflavin, and thiamin within 24 hours.

The test measures the extent to which the erythrocyte enzyme has been depleted of coenzyme, and the results are expressed either as the Activation coefficient or as the Percentage Stimulation:

$$\begin{aligned} \text{Activation coefficient} &= \frac{\text{activity of the coenzyme} - \text{stimulated enzyme}}{\text{activity of unstimulated enzyme}} \\ \text{Percentage Stimulation} &= \frac{\text{stimulated activity} - \text{basic activity}}{\text{basic activity}} \times 100\% \end{aligned} \tag{13.4.1}$$

Table 15.4 also presents comments on the three enzymes (transketolase, glutathione reductase, and aspartate aminotransferase) together with each of their corresponding vitamin-containing coenzymes (thiamine pyrophosphate, adenine dinucleotide, pyridoxal phosphate). Also given are values for activation coefficients for the three B vitamins (thiamine, riboflavin, and pyridoxine) and their interpretation. Ideally, the assay selected should: (a) reflect the amount of the nutrient available to the body, (b) respond rapidly to changes in supply of the nutrient, and (c) relate to the pathology of deficiency or excess. Measurement of the copper-containing enzyme lysyl oxidase is an example of an assay that fulfils these criteria. Connective tissue defects occur during the early stages of the copper deficiency syndrome. These defects can be attributed to the depressed activity of lysyl oxidase inhibiting cross-linking of collagen and elastin.

Table 15.4. erythrocyte stimulation tests of nutritional status for three vitamins.
AC: Activation coefficient
Modified from Bates CJ, Thurnham DI, Bingham SA, Margetts BM, Nelson M. (1997). Biochemical markers of nutrient status. In: Mar Press, Oxford, pp. 170–240.

Vitamin Thiamin	Enzyme Transketolase	Coe
Status: AC: 1.00-1.25	Normal or marginal status except when basic transketolase activity is low, then probably chronic deficiency.	Unst
Status: AC: > 1.25	Biochemical deficiency, high values likely to be acute deficiency	1.15 ·
Riboflavin	Glutathione reductase	ader
Status: AC: 1.00-1.30	Normal status	Very
Status: AC: 1.30-1.80	Marginal/deficient status	Mea
Status: AC: > 1.80	Deficient, intake <0.5mg riboflavin/d	Unre
Pyrodoxine	Aspartate aminotransferase	pyric
Status: AC: 1.00-1.50	Normal status	No a
Status: AC: 1.50-2.00	Marginal status	No a
Status: /AC: > 2.00	Deficient status	Unce

Many nutrients have more than one functional role and thus the activities of several enzymes may be affected during the development of a deficiency, thereby providing additional information on the severity of the deficiency state. For example, in the case of copper, platelet cytolysyl oxidasechrome c oxidase (Chapter 24) is more sensitive to deficiency than plasma erythrocyte superoxide dismutase, the activity of which is reduced only in more severe deficiency states (Milne and Nielsen, 1996).

15.4.3 Changes in blood components

Instead of measuring the activity of an enzyme, changes in blood components that are related to the intake of a nutrient can be measured. A well-known example is the measurement of hemoglobin concentrations in whole blood for iron deficiency anemia; iron is an essential component of the hemoglobin molecule (Chapter 17). Other examples include the determination of the two transport proteins — transferrin and retinol-binding protein (RBP) — as indicators of iron and vitamin A status, respectively, serum holotranscobalamin, a functional biomarker of vitamin B_{12} deficiency (Allen et al., 2018), and serum thyroglobulin, a thyroid-specific protein and a storage and synthesis site for thyroid hormones (Rohner et al., 2014;

Serum RBP is used increasingly as a proxy for serum retinol to assess vitamin A status at the population level, correlating closely with serum retinol concentrations, at least in individuals with normal kidney function who are not obese (Tanumihardjo et al., 2016). RBP in serum is also more stable, and easier and cheaper to analyze than retinol, although as with retinol, levels are reduced during inflammation. RBP is synthesized primarily in hepatocytes as the apo-form and secreted bound to retinol as the holo-RBP complex to provide vitamin A to peripheral tissues; one molecule of holo-RBP binds to one molecule of retinol. However, RBP is not secreted when stores of vitamin A are low and retinol limited. Because holo-RBP is complexed with transthyretin, loss of holo-RBP to glomerular filtration in the kidney is prevented.

Serum holotranscobalamin (holoTC), the component that delivers vitamin B_{12} to the tissues, has become increasingly used as a functional biomarker of B_{12}, with a specificity and sensitivity slightly higher than that of serum methylmalonic acid (MMA). Serum holoTC is most sensitive to recent intake, when concentrations can be increased even if stores are low. Concentrations of

holoTC, like serum MMA, are elevated in persons with impaired renal function, but are unaffected by pregnancy. Currently, there is no consensus on the cutoff to use and the assay of serum holoTC is expensive and not widely available (Allen et al., 2018). For more details see Chapter 22.

Serum thyroglobulin is recommended by WHO (2007) for monitoring the iodine status of school-aged children and a reference range for this age group has been established. Thyroglobulin concentrations in dried blood spots are also under investigation as a sensitive biomarker of iodine status in pregnant women (Stinca et al., 2017).

15.4.4 *In vitro* tests of *in vivo* functions

Tissue samples or cells can be removed from test subjects and isolated and maintained under physiological conditions. Attempts can then be made to replicate *in vivo* functions under *in vitro* conditions. Tests related to host-defense and immunocompetence are probably the most widely used assays of this type. They appear to provide a useful, functional, and quantitative measure of nutritional status.

Thymus-dependent lymphocytes originate in the thymus and are the main effectors of cell-mediated immunity. During protein-energy malnutrition, both the proportion and the absolute number of T-cells in the peripheral blood may be reduced.Peripheral T-lymphocytes are isolated from heparinized blood, then stained with fluorescent-labeled monoclonal antibodies (mABs), prior to analysis on a flow cytometer. The flow cytometer measures the properties of light scattering by the cells and the emission of light from fluorescent-labeled mAbs bound to the surface of the cell; details are given in Field (1996).

Lymphocyte proliferation assays are also examples of tests of this type. They are functional measures of cell-mediated immunity, assessed by the *in vitro* responses of lymphocytes to selected mitogens. Again, peripheral T-lymphocytes are isolated from blood and incubated *in vitro* with selected mitogens (Field, 1996). Details are summarized in Chapter 16.

Other *in vitro* tests include the erythrocyte hemolysis test and the dU suppression test, although the latter is no longer used. In the former, the rate of hemolysis of erythrocytes is measured; the rate correlates inversely with serum tocopherol levels (Chapter 18). Unfortunately, this test is not very specific, as other nutrients (e.g., selenium) influence the rate of erythrocyte hemolysis.

15.4.5 Load tests and induced responses *in vivo*

In the past, functional biomarkers conducted on the individual *in vivo* included load and tolerance tests (Solomons and Allen, 1983). Today, many of these tests are no longer used and have largely been superseded by other methods.

Load tests were used to assess deficiencies of water-soluble vitamins (e.g., tryptophan load test for pyridoxine, histidine load test for folic acid, vitamin C load test), and certain minerals (e.g., magnesium, zinc and selenium). In a load test, the baseline urinary excretion of the nutrient or metabolite is first determined on a timed preload urine collection (Robberecht and Deelstra, 1984). Then a loading dose of the nutrient or an associated compound is administered orally, intramuscularly, or intravenously. After the load, a timed sample of the urine is collected and the excretion level of the nutrient or a metabolite determined. The net retention of the nutrient is calculated by comparing the basal excretion data with net excretion after the load. In a deficiency state, when tissues are not saturated with the nutrient, excretion of the nutrient or a metabolite will be low because net retention is high.

The relative dose response (RDR) test is the most well known functional *in vivo* load test in use today. This test is accepted as a functional reference method to assess the presence or absence of low vitamin A stores in the liver (Chapter 18). However, in the RDR test, unlike the conventional loading tests described above, the response is greatest in deficient individuals. The principal of RDR is based on the observation that in vitamin A inadequacy, retinol binding protein (RBP) that has not bound to retinol (apo-RBP) accumulates in the liver. Following the administration of a test dose of vitamin A (commonly in the form of retinyl palmitate), some of the retinol binds to the accumulated apo-RBP in the liver and the resulting holo-RBP (i.e., RBP bound to retinol) is rapidly mobilized from the liver into the circulation. In individuals with vitamin A deficiency, a small dose of retinyl palmitate leads to a rapid sustained increase in serum retinol, whereas in vitamin A replete individuals, there is very little increase. An RDR value > 20% is considered to reflect vitamin A stores of < 0.7µmol/g liver (WHO, 1996). Investigations are underway to explore the assessment of the RDR test based on serum RBP to determine low hepatic vitamin A stores instead of serum retinol in an effort to eliminate the need to use HPLC for the serum retinol assay (Fujita et al., 2009).

The modified relative dose response (MRDR) has been developed as an alternative because the RDR test requires two blood samples per individual. The MRDR uses 3,4-didehydroretinyl acetate (DRA), or vitamin A_2 instead of retinyl palmitate as the challenge dose, and requires only a single blood sample, taken between 4 and 7h after dosing. Serum is analyzed for both 3,4,-didehydroretinol (DR) and retinol in the same sample, and the ratio of DR to retinol in serum is called the MRDR value and used to indicate liver reserves. Values ≥ 0.060 at the individual level usually indicate insufficient liver reserves (≤ 0.1μmol retinol/g), whereas values < 0.06 are indicative of sufficient liver reserves (≥ 0.1μmol retinol/g). Group mean ratios of < 0.030 appear to correlate with adequate status. The MRDR test has been used in numerous population groups, in both children and adults worldwide. The values obtained, however, are not useful for defining vitamin A status above adequacy. For more details of the use of RDR and MRDR as functional biomarkers of vitamin A status, see Tanumihardjo et al. (2016).

The qualitative CobaSorb test is another example of an *in vivo* load test. It is used to detect malabsorption of vitamin B_{12} and has replaced the earlier Schilling test and its food-based version (using cobalamin-labeled egg yolk), which have been discontinued. For the CobaSorb test, a dose of 9mg of crystalline B_{12} in water is administered orally at 6h intervals over a 24h period, and the increase in serum holo-transcobalamin measured on the following day. The test is a qualitative assay and is used to determine if patients will respond to low-dose B_{12} supplements or will require treatment with pharmacological doses. The test does not provide a quantitative estimate of bioavailability of vitamin B_{12} — for more details of the test, see Brito et al.(2018).

Delayed-type hypersensitivity is a well known example of a biomarker based on an induced response *in vivo*. This is a direct functional measure of cell-mediated immunity used in both hospital and community settings. Suppression of cell-mediated immunity signals a failure of multiple components of the host-defense system. The test involves injecting a battery of specific antigens intradermally into the forearm; those commonly used are purified protein derivative (PPD), mumps, Tricophyton, *Candida albicans*, and dinitrochlorobenzene (DNCB). In healthy persons re-exposed to recall antigens intradermally, the recall antigens induce the T-cells to respond first by proliferation and then by the release of soluble mediators of inflammation, producing an induration (hardening) and erythema (redness). This induced response is noted at selected time intervals, and is often reduced in persons with protein-energy malnutrition and micronutrient deficiencies such as vitamin A, zinc, iron and pyridoxine. However, the test is not specific enough to detect individual micronutrient deficiencies (Raiten et al., 2015). For details of the technique, interpretation, and some of the limitations of DTH skin testing, see Ahmed and Blose (1983).

15.4.6 Spontaneous *in vivo* responses

Functional tests based on spontaneous *in vivo* responses often measure the net effects of contextural factors that may include social and environmental factors as well as nutrition. Hence, they are less sensitive and specific than biomarkers that assess nutrient exposure, status, or biochemical function (Raiten and Combs, 2015). As a consequence, they should be assessed alongside more specific biomarkers so that the functional impact of the status of a specific nutrient can be identified.

Formal dark adaptometry is one of several functional biomarkers based on spontaneous physiological *in vivo* responses that exist for vitamin A. It was the classical method for assessing night blindness (i.e., poor vision in low-intensity light) associated with vitamin A deficiency. This condition arises when the ability of the rod cells in the retina to adapt in the dark, and the ability of the pupils to properly meter light in and out of the eye, are impaired. However, the equipment used for formal dark adaptometry was cumbersome, and the method very time-consuming, so it is no longer used.

Rapid dark adaptation test (RDAT) has superseded formal dark adaptometry, with results that correlate with those of the classical method. The test is based on the measurements of the timing of the Purkinje shift, in which the peak wavelength sensitivity of the retina shifts from the red toward the blue end of the visual spectrum during the transition from photopic or cone-mediated day vision to scotopic or rod-mediated night vision. This shift causes the sensitivity of blue light to appear brighter than that of red light under scotopic lighting conditions. The test requires a light-proof room, a light source, a dark, non-reflective work surface, a standard X-ray view box, and sets of red, blue, and white discs; details are given in Vinton and Russell (1981). The RDAT, however, is not appropriate for young children.

Pupillary threshold test can be used for children from age 3y, for adults, and under field conditions (Tanumihardjo et al., 2016). The test measures the threshold of light at which pupillary contraction occurs under dark adapted conditions. Minimal cooperation from the subjects is required for the test which is performed in a darkened facility, often a portable tent, and takes about 20min per subject. Special pairs of goggles have been invented to measure the pupillary response to light stimuli — details are given in Chapter 18.

Capillary fragility has been used as a functional biomarker of vitamin C deficiency since 1913 because frank petechial

hemorrhages occur in overt vitamin C deficiency. The test, however, is not very specific to vitamin C deficiency states (see Chapter 19); static biochemical tests are preferred to assess the status of vitamin C.

Taste acuity can be associated with suboptimal zinc status in children and adults, (Gibson et al., 1989), as well as in some disease states in which secondary zinc deficiency may occur (e.g., cystic fibrosis, Crohn's disease, celiac sprue and chronic renal disease (Desor and Maller, 1975; Kim et al., 2016). Positive associations between taste acuity for salt and biomarkers of zinc status (i.e., erythrocyte zinc) have been reported in the elderly(Stewart-Knox et al., 2005). Moreover, an increase in taste acuity for salt was reported in older adults in response to 30mg zinc/day compared to a placebo during a six-month double-blind randomized controlled trial (Stewart-Knox et al., 2008).

Taste acuity can be assessed by using the forced drop method that measures both the detection and recognition thresholds (Buzina et al., 1980), or the recognition thresholds only (Desor and Maller, 1975). An electrogustometer, which measures taste threshold by applying a weak electric current to the tongue, has been used in some studies (Prosser et al., 2010). Many other factors affect taste function, and taste acuity alone should not be used to measure zinc status.

Handgrip strength measured using a dynamometer, is a well-validated proxy measurement for lower-body strength (Abizanda et al., 2012). It has been used in several intervention studies designed to improve muscle strength and function among non-malnourished sarcopenic older adults at high risk for disability (Bauer et al., 2015; Tieland et al., 2015).

15.4.7 Growth responses

Responses in growth have limited sensitivity and are not specific for any particular nutrient, and hence are preferably measured in association with other more specific nutrient biomarkers.

Linear growth is considered the best functional biomarker associated with the risk of zinc deficiency in populations. It is is usually measured alongside serum zinc, a biomarker of zinc exposure and status at the population level. The International Zinc Consultative Group (IZiNCG) recommend using the percentage of children under five years of age with length-for-age (LAZ) or height-for-age z-score (HAZ) < –2 as a functional biomarker to estimate the risk of zinc deficiency in a population (Technical Brief No.01, 2007; de Benoist et al., 2007).

The WHO Global Database on Child Growth and Malnutrition also use LAZ or HAZ < –2 to define children as “stunted”, and include stunting as one of the six global nutrition targets for 2030. In a healthy population, about 2.5% of all children have a LAZ or HAZ < –2. In communities where short stature is the norm, stunting often goes unrecognized.

Figure 15.10 presents the distribution of length/height-for-age Z-scores of children from the India National Family Health Survey 2005–2006, and shows the entire distribution shifted to the left compared with the WHO Child Growth Standards (WHO, 2006). These results highlight the fact that those children who are stunted are only a subset of those with linear growth retardation. Here all the children were affected by some degree of linear growth retardation (de Onis and Branca, 2016).

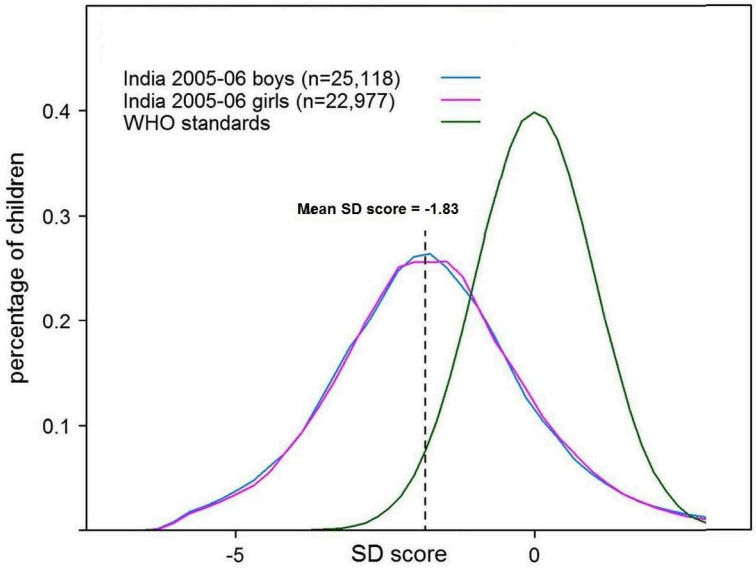


Figure 15.10. Distribution of length/height-for-age Z-scores of children from the Indian National Family Health Survey 2005–2006. Modified from de Onis and Branca (2016).

Height-for-age difference (HAD) defined as a child's height minus the median reference value of height-for-age of WHO Child Growth Standard, expressed in centimeters, is recommended to describe and compare changes in height as children age (Leroy et al., 2014). Leroy and colleagues argue that HAZ is inappropriate to evaluate changes in height as children age because HAZ scores are constructed using standard deviations from cross-sectional data that change with age. Leroy et al. (2015) compared changes in growth in populations of children 2–5y using HAD vs. HAZ from cross-sectional data based on six Demographic and Health Surveys (DHS). There was no evidence of population-level catch-up in linear growth in children aged 2–5y when using HAD, but instead a continued deterioration reflected in a decrease in mean HAD between 2 and 5y. In contrast, based on HAZ, there was no change in mean HAZ (Leroy et al., 2015); see Chapter 13 for more details.

Linear growth velocity is also used as a functional biomarker of malnutrition in infants and young children. It can be assessed via measurements of changes in recumbent length for children < 2y and changes in height for older children. A high degree of precision is required for these measurements because two measurements are needed. During infancy, length increments can be assessed at 1mo intervals for the first 6mos and at 2mos intervals from 6 to 12mos. Increments measured over 6mos are the minimum interval that can be used to provide reliable data during adolescence (WHO, 1995). Seasonal variation in growth may occur. In high-income countries, height velocity, for example, may be faster in the spring than in the fall and winter.

The following formula is used to calculate velocity: Velocity = $(x_2 - x_1) / (t_2 - t_1)$ where x_2 and x_1 are values of the measurement on two occasions t_2 and t_1. Length or height velocity is normally expressed as cm/y. Currently, no uniform criteria exist for defining growth faltering based on growth velocity data, although in practice zero growth in two consecutive periods is sometimes used (WHO, 1995). WHO has developed a set of growth velocity charts which are recommended for international use (de Onis et al., 2011).

Knee height measurements using a portable knemomter can be used to provide a more sensitive short-term measure of growth velocity in children > 3y based on lower leg length (Davies et al., 1996). This equipment was used to obtain accurate measurements of the lower leg length of indigenous Shuar children aged 5–12y from Ecuador; the technical error of the measurement (TEM) was low — 0.18mm (Urlacher et al., 2016). For children < 3y, a mini-knemometer can be used to measure lower leg length (Kaempf et al., 1998).

15.4.8 Developmental responses

The assessment of cognitive function requires rigorous methodology. Even with careful methodology, a relationship between cognitive function and nutrient deficiency can be established only by: (a) documenting clinically important differences in cognitive function between deficient subjects and healthy placebo controls, and (b) demonstrating improvement in cognitive function after

an intervention. The individuals should be matched, and the design should preferably be a double-blind, placebo-controlled, randomized intervention (Lozoff and Brittenham, 1986).

To date, for example, results of meta-analyses have concluded that there is no clear evidence of the benefits of iron supplementation on visual, cognitive, or psychomotor development in preschool children (Larson and Yousafzai., 2017). In contrast, evidence for the benefits of such supplementation on cognitive performance for school-aged children who are anemic at baseline is strong (Low et al., 2013). Larson et al. (2017) have emphasized the need to conduct high-quality placebo-controlled, adequately powered trials of iron interventions on cognitive performance in young children to resolve the current uncertainties.

Several measurement scales of cognitive function are available, some of which are summarized briefly below.

Bayley Scales of Infant and Toddler Development are the most widely used method worldwide for assessing various domains of cognitive function in infants and toddlers (Albers and Grieve, 2007). The third edition of the scales (Bayley-III) measures child development across five domains: cognition, receptive and expressive language, motor, adaptive, and social-emotional skills. They were constructed in the U.S. and have norms based on an American sample, so cultural adaptions are often needed when using them elsewhere. Appropriate training and standardization are a prerequisite to obtain reliable assessment across testers.

Ages and Stages questionnaire (ASQ-3), a parental screening tool, is frequently used in large-scale research studies, as it is a cheaper, and a less time-consuming measure of early childhood development. The ASQ-3 consists of 30 simple, straightforward questions covering five skillsets of childhood development: problem solving, communication, fine motor skills, gross motor skills, and personal social behavior. The ASQ-3 was also developed in the U.S. to identify infants and toddlers 1–66mos at risk of a developmental delay (Steenis et al., 2015).

Several studies have compared the ASQ-3 with the Bayley scales to assess developmental level of infants and toddlers. Substantial variations in the sensitivity and specificity of the ASQ-3 across studies have been reported. Such variations may be due in part due to differences in study design, study samples (high or low risk), versions of the Bayley scales used, as well as the countries in which the studies were conducted. In several of these comparative studies, the ASQ-3 has been found to have only low to moderate sensitivity (Steenis et al., 2015; Yue et al., 2019).

Fagan Test of Infant Intelligence is also used to assess cognitive function at four ages: 27,29,39, and 52 weeks postnatal age corrected for prematurity. The test is made up of 10 novelty problems, which comprise one familiar and one novel stimulus presented simultaneously. All the stimuli are pictures of faces of infants, women, and men. A novelty preference score for each age is calculated as the average percent of time spent fixating the novel picture across the 10 problems (Andersson, 1996). This test was used in a large trial in which Chilean infants 6–12mos (n=1123) were supplemented with iron and compared to a no-added-iron group (n=534) (Lozoff et al., 2003).

Wechsler Preschool and Primary Scale of Intelligence (WISC) is designed for children age 6y through 16y 11mos. WISC-IV contains 10 core subsets and five supplementary subtests. The core subsets consist of Block Design, Similarities, Digit Span, Reasoning, Coding, Vocabulary, Letter-Number Sequencing, Symbol Search, Comprehension, and Picture Concepts. The five supplementary subtests comprise information, Word Reasoning, Picture Completion, Arithmetic, and Cancellation. The ten core subsets combine to form four composite index scores: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI). The Full Scale Intelligence Quotient (FSIQ) is derived from the sum of the 10 core subtest scores (Watkins and Smith, 2013). The WISC-IV has now been modified as the WISC-V, which can be administered more quickly than previously, and with more accurate scoring, electronically (Na and Burns, 2016). The WISC-III was used in a follow-up study to investigate the effect of folic acid supplementation during trimesters 2 and 3 of pregnancy on cognitive performance in the child at 7y (McNulty et al., 2019).

Raven's Progressive Matrices (RPM) and Raven's Mill Hill Vocabulary Scales (MHV) have been used in school children and adults to assess basic cognitive functioning. The RPM test is made up of a series of diagrams or designs with a part missing. Respondents are asked to select the correct part to complete the designs from a number of options printed beneath. The MHV scale consists of 88 words, arranged in order of ascending difficulty which respondents are asked to define. Both the RPM and MHV have been used for different cultural, ethnic, and socioeconomic groups worldwide(Raven, 2000).

Raven's Progressive Matrices were used in a study of Kenyan school children designed to test whether animal source foods have a key role in the optimal cognitive development of children. Results are shown in Figure 15.11. Post-hoc analyses showed that children who received a supplement with meat had significantly greater gains on the Raven's Progressive Matrices than any other group (Whaley et al., 2003).

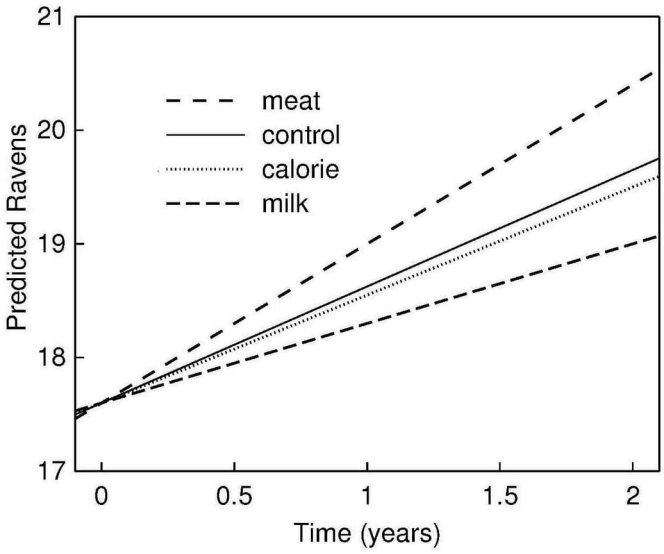


Figure 15.11. Longitudinal regression curves for Raven's scores across five time points, indicating improved cognition in some Kenyan children under four different dietary regimes. Data from Whaley et al., 2003, by permission of Oxford University Press.

Mini-Mental State Examination (MMSE) is the most studied instrument for use as a screening measure of cognitive impairment in the elderly (Lin et al., 2013). The MMSE is divided into two parts and is not timed. The first part requires vocal responses only and covers orientation, memory, and attention, with a maximum score of 21. The second part tests the ability to name, follow verbal and written commands, write a senetence spontaneously, and copy a complex polygon. This part has a maximum score of 9, giving a maximum total score of 30; for more details see Folstein and Folstein (1975).

The MMSE was used to assess cognitive decline at six monthly intervals in a 3y double-blind, placebo controlled randomized controlled trial of healthy postmenopausal African American women aged 65y and older (n=260). Women were randomized to receive vitamin D (adjusted to achieve a serum level > 30ng/mL) with calcium (diet and supplement total of 1,200mg/d) or placebo (with calcium supplement of 1,200mg/d). Over three years there was no difference in cognition between the two groups, providing no support in this trial for providing a vitamin D intake greater than the recommended daily allowance for the prevention of cognitive decline (Owusu et al., 2019).

Other instruments such as the Clock Drawing test (CDT), Mini-Cog, Memory Impairment Screen, Abbreviated Mental Test (AMT), and Short Portable Mental Status Questionnaire (SPMSQ) can be used to detect dementia, although with more limited evidence. Moreover, for the AMT and SPMSQ, evidence of their usefulness in English is limited (Lin et al., 2013).

Motor development is also an essential component of child development. In several randomized controlled trials in low income countries, positive effects on gross motor milestones, particularly attainment of walking unassisted have been reported in infants receiving iron and/or zinc supplements or micronutrient-fortified-complementary foods (Adu-afarwuah et al., 2007; Black et al., 2004; Bentley et al., 1997 ; Masude and Chitundu, 2019).

Readers are advised to consult the WHO website for details of the The Motor Development Study undertaken as a component of the WHO Multicenter Growth Reference Study. (MGRS). During this longitudinal study, data were collected from five countries (Ghana, India, Norway, Oman, and the United States) on six gross motor milestones using standardized testing procedures. The gross motor milestones and their performance criteria are outlined in Box 15.9: see Wijnhoven et al. (2004). for further details on the methods and training and standardization of fieldworkers. Because achievements of the six milestones were assessed repeatedly between 4–24mos, the sequence and tempo of the milestones as well as the ages of their attainment can be documented (Wijnhoven et al., 2004). Use of these WHO data are recommended for future studies involving assessment of gross

motor development.

Box 15.9. MGRS

- **Sitting without support**: Child sits up straight with the head erect for at least 10 seconds. Child does not use arms or hands to balance body or support position.
- **Hands-and-knees crawling**: Child alternately moves forward or backward on hands and knees. The stomach does not touch the supporting surface. There are continuous and consecutive movements, at least three in a row.
- **Standing with assistance**: Child stands in upright position on both feet, holding onto a stable object (e.g., furniture) with both hands without leaning on it. The body does not touch the stable object, and the legs support most of the body weight. Child thus stands with assistance for at least 10 seconds.
- **Walking with assistance**: Child is in upright position with the back straight. Child makes sideways or forward steps by holding onto a stable object (e.g., furniture) with one or both hands. One leg moves forward while the other supports part of the body weight. Child takes at least five steps in this manner.
- **Standing alone**: Child stands in upright position on both feet (not on the toes) with the back straight. The legs support 100% of the child's weight. There is no contact with a person or object. Child stands alone for at least 10 seconds.
- **Walking alone**: Child takes at least five steps independently in upright position with the back straight. One leg moves forward while the other supports most of the body weight. There is no contact with a person or object.

15.4.9 Depression

Links between depression and micronutrient deficiencies have been reported for folate, vitamin B_{12}, calcium, magnesium, iron, selenium, zinc, and n-3 fatty acids. Several mechanisms have been proposed, including mitochondrial function (including inadequate energy), disturbances in normal metabolism, genetic polymorphisms requiring increased or atypical nutrient requirements, increased inflammation, oxidative stress, and alterations in the microbiome (Campisi et al., 2020). Of the micronutrients, those most frequently studied have been zinc, vitamin D, iron, folate, and vitamin B_{12} in investigations in children, adolescents and the elderly. There have been reports of improvements in patients with depression given supplemental zinc, especially when the supplemental zinc is used as an adjunct to conventional antidepressant drug therapy (Ranjbar et al., 2014). However, methodological limitations exist in some of the studies, especially those on children and adolescents, and more well-designed and adequately powered placebo-controlled randomized controlled trials are needed (Campisi et al., 2020).

Beck Depression Inventory-II (BDI-II) is frequently used to measure depression (Richter et al., 1998; Levis et al., 2019). This instrument is said to be a cost-effective questionnaire with a high reliability and capacity to discriminate between depressed and non-depressed individuals and which is applicable to both research and clinical practice worldwide (Wang and Gorenstien, 2013).

Patient Health Questionnaire (PHQ) is a useful screening tool to detect major depression in population-based studies. Depression can be defined with PHQ-9 by using a cutoff point of 10 or above regardless of age, although specificity of the PHQ-9 may be less for younger than for older patients (Levis et al., 2019).

This page titled 13.4: Biomarkers of function (15.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13.5: Factors affecting choice of nutritional biomarkers (15.5)

Biomarkers should be selected with care, and their limitations under conditions of health, inflammation, genetic and disease states understood. Several biological factors must be taken into account when selecting nutritional biomarkers to assess nutritional status and these are discussed more fully in the following sections. They are also affected by non-biological sources of variation arising from specimen collection and storage, seasonality, time of day, contamination, stability, and laboratory quality assurance. Both the biological and non-biological sources of variation will impact on the validity, precision, accuracy, specificity, sensitivity, and predictive value of the biomarker. Because almost all techniques are subject to both random and systematic measurement errors, personnel should use calibrated equipment and should be trained to use standardized and validated techniques which are monitored continuously by appropriate quality-control procedures.

15.5.1 Study objectives

The choice of nutritional biomarkers is strongly influenced by the study objectives. Nutritional biomarkers can be used to determine the health impacts of nutritional status at the population and/or the individual level. At the population level, factors such as cost, technical and personnel requirements, feasibility, and respondent burden are important considerations when choosing biomarkers. Population-level assessment is used to develop programs such as surveillance, to identify populations or sub-groups at risk, to monitor and evaluate public health programs, and to develop evidence-based national or global policies related to food, nutrition, and health.

Nutritional biomarkers are also used at the individual level by clinicians to assess the nutritional status of patients who are “apparently healthy” or “apparently sick”, or who have subclinical illnesses. They may also be used to predict the future risk of disease or long-term functional outcomes if abnormal values persist, and to generate data to support evidence-based clinical guidelines (Combs et al., 2013). The effects of genetic polymorphisms on the clinical usefulness of biomarkers are increasingly being recognized; see Section 15.7.1. for more details.

Whether the study is at the population or individual level can influence, for example, the choice of a biomarker of nutritional exposure. To determine the risk of nutrient inadequacy in a national survey, a single 24hr recall per person (with repeats on at least a subsample) is required so that the distribution of usual intakes can be adjusted statistically. To assess the usual dietary intakes of a patient for dietary counseling, a food frequency questionnaire or dietary history is often used. For more details of these dietary assessment methods, see Chapter 3.

15.5.2 Population and setting

Factors such as life-stage group and ethnicity must also be taken into account when selecting nutritional biomarkers. In studies during early infancy, local or national ethics committees may prohibit the collection of venipuncture blood samples, and instead suggest that less invasive biomarkers based on urine, saliva, hair, or fingernails are used. Hormonal changes during pregnancy, along with an increase in plasma volume during the second and third trimester affect concentrations of several micronutrient biomarkers (e.g., serum zinc and vitamin B_{12}), making it essential to identify women who are pregnant in order to ensure the correct interpretation of biomarker values.

Rural settings may present many challenges associated with the appropriate collection, transport, centrifugation and storage of biomarkers; for example, it may be difficult to ensure a temperature-controlled supply chain or “cold chain” for specimen collection. Increasingly, serum retinol binding protein (RBP) is being used as a surrogate for serum retinol in studies at the population level. As noted earlier, serum RBP is more stable, and the assay is easier, cheaper, and correlates closely with serum retinol, provided that the individuals tested are neither obese nor have abnormal kidney function (Tanumihardjo et al., 2016).

15.5.3 Validity

Validity refers to how well the biomarker correctly describes the nutritional parameter of interest. As an example, if the biomarker selected reflects recent dietary exposure, but the study objective is to assess the total body store of a nutrient, the biomarker is said to be invalid. In U.S. NHANES I, thiamin and riboflavin were analyzed on casual urine samples because it was not practical to collect 24h urine specimens. However, results were not indicative of body stores of thiamin or riboflavin, and hence were con-

sidered invalid and thus not included in U.S. NHANES II or U.S. NHANES III(Gunter and McQuillan, 1990). Valid biomarkers are ideally free from random and systematic errors and are both sensitive and specific. Unfortunately the action of inflammation, stress, or certain medications,on enzyme activity and nutrient metabolism may alter nutrient status and thus affect the validity of a nutritional biomarker. As an example, the acute-phase response observed during an infectious illness may cause changes in certain nutrient levels in the blood (e.g., plasma zinc may fall and retinol binding protein fall, whereas plasma ferritin increases), that do not reflect alterations in the nutrient status *per se*, but indicate instead a redistribution of the nutrient mediated by the release of cytokines (Raiten et al., 2015).

In view of the possible effects of inflammation on biomarker levels, measures of infection status should be assessed concurrently. For example, to adjust for the presence of systemic inflammation, WHO has recommended the concurrent measurement of two inflammatory biomarkers — serum C-reactive protein and α-1-acid glycoprotein — so that an adjustment using regression modeling can be applied (Suchdev et al., 2016).

Table 15.5 shows, for Indonesian infants at 12mos, the impact of inflammation on the geometric mean and prevalence estimates of iron, vitamin A, and zinc deficiency based on serum ferritin, RBP, and zinc. Note the decrease in geometric mean for serum ferritin but the corresponding increases for serum RBP and zinc after applying the recommended BRINDA adjustment for inflammation. As a consequence, there is a marked increase in the estimated proportion at risk to low serum ferritin (indicative of depleted iron stores), and a marked decrease in the estimate of prevalence of both vitamin A deficiency and zinc deficiency Table 15.5

Table 15.5. Impact of inflammation on micronutrient biomarkers of Indonesian infants of age 12mos. From Diana et al.(2018).
** Ferritin < 12µg/L*
*** RBP < 0.83µmol/L*
**** Zinc < 9.9µmol/L*

Biomarker in serum	Geometric mean (95% CI)	Proportion at ri
Ferritin*: No adjustment	14.5 µg/L (13.6–17.5)	44.9
Ferritin: Brinda adjustment	8.8 µg/L (8.0–9.8)	64.9
Retinol binding protein**: No adjustment	0.98 (µmol/L) (0.94–1.01)	24.3
Retinol binding protein: Brinda adjustment	1.07 µmol/L (1.04–1.10)	12.4
Zinc***: No adjustment	11.5 µmol/L (11.2–11.7)	13.0
Zinc: Brinda adjustment	11.7 µmol/L (11.4–12.0)	10.4

Other disease processes may alter the nutrient status as a result of impaired absorption, excretion, transport, or conversion to the active metabolite and thus confound the validity of the chosen biomarker. In some cases the cause of these disease processes is hereditary, but in other cases the cause is acquired. Some examples of disease processes that affect nutrient status and, in turn, nutritional biomarkers, are shown in Table 15.6.

Table 15.6. Examples of some disease states that may confound the validity of laboratory tests. In: Biochemical markers of nutrient s

Disease	Biomarkers of nutrient indices that may
Pernicious anemia	Vitamin B_{12} (secondary effect on folate)
Vitamin-responsive metabolic errors	Usually B-vitamins (e.g., vitamins B_{12}, B_6, r
Tropical sprue	Vitamins B_{12} and folate (local deficiencies
Steatorrhea	Fat-soluble vitamins, lipid levels, energy
Abetalipoproteinemia	Vitamin E
Thyroid abnormality	Riboflavin, iodine, selenium, lipid levels, er
Diabetes	Possibly vitamin C, zinc, chromium, and se

Disease	Biomarkers of nutrient indices that may
Infections, inflammation, acute phase reaction	Zinc, copper, iron, vitamin C, vitamin A, lip
Upper respiratory tract infections, diarrheal disease, measles	Especially vitamin A, lipid levels, protein
Renal disease	Increased retention or increased loss of m
Cystic fibrosis	Especially vitamin A, lipid levels, protein
Various cancers	Lowering of vitamin indices
Acute myocardial infarction	Lipid levels affected for about 3 mo
Malaria, hemolytic disease, hookworm, etc	Iron, vitamin A, lipid
Huntington's chorea	Energy
Acrodermatitis enteropathica; various bowel, pancreatic, and liver diseases	Zinc, lipid levels, protein
Hormone imbalances	Minerals, corticoids, parathyroid hormone, agents and estrogen therapy

Depending on the biomarker, potential interactions with several physiological factors such as fasting status, diurnal variation, time of previous meal consumption and homeostatic regulation must also be considered. For instance, fluctuations in serum zinc in response to meal consumption can be as much as 20% (King et al., 2015).

15.5.4 Precision

Precision refers to the degree to which repeated measurements of the same biomarker give the same value. The precision of a nutritional biomarker is assessed by repeated measurements on a single specimen or individual. The coefficient of variation (CV), as determined by the ratio of the standard deviation to the mean of the replicates (SD/mean × 100%) is the best quantitative measure of the precision. Ideally, the CV should be calculated for specimens at the bottom, middle, and top of the reference concentration range for the biomarker, as determined on apparently healthy individuals. These same specimens then serve as quality controls.

Typically, the quality-control specimens used to calculate the CV are pooled samples from donors similar to the study participants. It is important that these quality-control specimens should, to the analyst, appear identical to the specimens from the study participants. This means that the same volume, type of vial, label and so on should be used.

Quality-control specimens should be inserted blind into each batch of specimens from the study participants. Both the intra- and inter-run CVs should be calculated on these quality-control specimens. The former is calculated from the values for aliquots of the quality-control specimens analyzed within the same batch, and the latter normally calculated from the values for aliquots of the quality-control specimens analyzed on different days (Blanck et al., 2003).

The precision of the measurement of a biomarker is in part a function of the random measurement errors that occur during the actual analytical process, and in some cases also a function of the intra-individual biological variations that occur naturally over time. The relative importance of these two sources of uncertainly varies with the measurement. For some biochemical measurements (e.g., serum iron), the intra-individual biological variation is quite large: coefficients of variation may exceed 30%, and be greater than any analytical variation. Consideration of intra-individual variation is also important when assessing dietary exposure, because nutrient intakes of an individual always vary over time. However, in this case, the intra-individual variation is a measure of the “true day-to-day” variation in the dietary intake of an individual. Strategies exist to account for the impact of intra-individual variation on the measurement of true usual intake of an individual — see chapter 6 for more details.

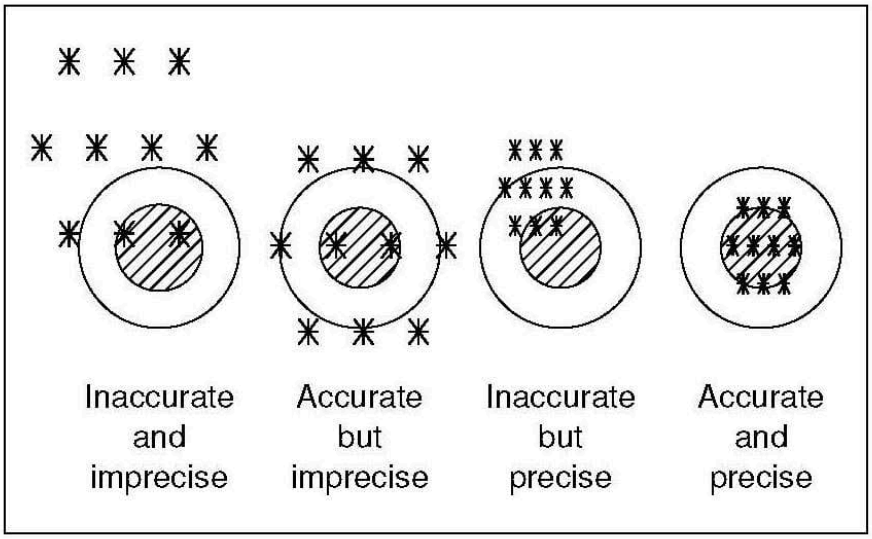


Figure 15.12. Differences between precision and accuracy.

The *attainable* level of precision for the measurement of any particular biomarker depends on the procedure, whereas the*required* precision is a function of the study objectives. Some investigators have stipulated that, ideally, the analytical CV for an assay used in epidemiological studies should not exceed 5%. In practice, this level of precision is difficult to achieve for many assays and less precise measurements in epidemiological studies may result in a failure to detect a real relationship of the nutritional biomarker and the outcome of interest (Blanck et al., 2003). Of note, as shown in Figure 15.12, even if the precision is acceptable, the analytical method may not be accurate.

15.5.5 Sensitivity and specificity

Sensitivity refers to the extent to which the biomarker identifies individuals who genuinely have the condition under investigation (e.g., a nutrient deficiency state). Sensitive biomarkers show large changes as a result of only small changes in nutritional status. A biomarker with 100% sensitivity correctly identifies all those individuals who are genuinely deficient; no individuals with the nutrient deficiency are classified as "well" (i.e., there are no false negatives). Numerically, sensitivity is the proportion of individuals with the condition who have positive tests (true positives) divided by the sum of true positive and false negatives. The sensitivity of a biomarker changes with the prevalence of the condition as well as with the cutoff point.

Biomarkers that are strictly homeostatically controlled have very poor sensitivity. Figure 15.1 shows the relationship between mean plasma vitamin A and liver vitamin A concentrations. Note that plasma retinol concentrations reflect the vitamin A status only when liver vitamin A stores are severely depleted (< 0.07µmol/g liver) or excessively high (> 1.05µmol/g liver). When liver vitamin A concentrations are between these limits, plasma retinol concentrations are homeostatically controlled and levels remain relatively constant and do not reflect total body reserves of vitamin A. Hence, in populations from higher income countries where liver vitamin A concentrations are generally within these limits, the usefulness of plasma retinol as a sensitive biomarker of vitamin A exposure and status is limited (Tanumihardjo et al., 2016). Likewise, the use of serum zinc as a biomarker of exposure or status at the individual level is limited due to tight homeostatic control mechanisms. Based on a recent meta-analysis, doubling the intake of zinc was shown to increase plasma zinc concentrations by only 6% (King, 2018).

Specificity refers to the ability of a nutritional biomarker to identify and classify those persons who are genuinely well nourished. If the biomarker has 100% specificity, all genuinely well-nourished individuals will be correctly identified; no well-nourished individuals will be classified as under-nourished (i.e., there are no false positives). Numerically, specificity is the proportion of individuals without the condition who have negative tests (true negatives divided by the sum of true negatives and false positives).

Unfortunately, many of the health and biological factors noted in Box 15.2 and diseases summarized in Table 15.6 reduce the specificity of a biomarker. Inflammation, for example, reduces serum zinc (Table 15.5), yielding a concentration that does not reflect true zinc status, so misclassification occurs; individuals are designated "at risk" with low serum zinc concentrations when they are actually unaffected (false positives). In contrast, inflammation increases serum ferritin, so that in this case individuals

may be designated “not at risk” when they are truly affected by the condition (false negatives).

The ideal biomarker has a low number of both false positives (high specificity) and false negatives (high sensitivity), and hence is able to completely separate those who genuinely have the condition from those individuals who are healthy. In practice, a balance has to be struck between specificity and sensitivity, depending on the consequences of identifying false negatives and false positives.

15.5.6 Analytical sensitivity and analytical specificity

Unfortunately, the term “sensitivity” is also used to describe the ability of an *analytical* method to detect the substance of interest. The more specific term “analytical sensitivity” should be used in this context.

For any analytical method, the smallest concentration that can be distinguished from the blank is termed the “analytical sensitivity” or the “minimum detection limit.” The blank should have the same matrix as the test sample and, therefore, usually contains all the reagents but none of the added analyte. Recognition of the analytical sensitivity of a biochemical test is particularly important when the nutrient is present in low concentrations (e.g., the ultra-trace elements Cr, Mn, and Ni).

In practical terms, the minimum detection limit or the analytical sensitivity is best defined as three times the standard deviation (SD) of the measurement at the blank value. To calculate the SD of the blank value, 20 replicate measurements are generally recommended. Routine work should not include making measurements close to the detection limit and should normally involve analyzing the nutrient of interest at levels at least five times greater than the detection limit. Measured values at or below the detection limit should not be reported.

The ability of an analytical method to measure exclusively the substance of interest is a characteristic referred to as the “analytical specificity.” Methods that are nonspecific generate false-positive results because of interferences. For example, in U.S. NHANES II, the radioassay used gave falsely elevated results for vitamin B_{12}. This arose because the porcine intrinsic factor (IF) antibody source initially used reacted both with vitamin B_{12} and with nonspecific cobalamins present in serum. As a result, erroneously high concentrations were reported and the samples had to be reanalyzed using a modified method based on purified human IF, specific for vitamin B_{12} (Gunter and McQuillan, 1990).

Strategies exist to enhance analytical specificity (and sensitivity). Examples include the use of dry ashing or wet digestion to remove organic material prior to the analysis of minerals and trace elements.

15.5.7 Analytical accuracy

The difference between the reported and the true amount of the nutrient/metabolite present in the sample is a measure of the analytical accuracy (“trueness”) of the laboratory test (Figure 15.12). Guidelines on choosing a laboratory for assessment of a nutritional biomarker are given in Blanck et al. (2003).

Several strategies can be used to ensure that analytical methods are accurate. For methods involving direct analysis of nutrients in tissues or fluids, a recovery test is generally performed. This involves the addition of known amounts of nutrient to the sample. These spiked samples are then analyzed together with unspiked aliquots to assess whether the analytical value accounts for close to 100% of the added nutrient.

As an additional test for accuracy, aliquots of a reference material, similar to the sample and certified for the nutrient of interest, should be included routinely with each batch of specimens. If possible, several reference materials, with values spanning the range observed in the study samples, should be analyzed (Blanck et al., 2003). Such a practice will document the accuracy achieved.

Standard reference materials (SRMs) can be obtained from the U.S. National Institute of Standards and Technology (NIST) (for serum Zn, vitamins $B_6$and B_{12}, folate, vitamin D, carotenoids), the U.S. Centers for Disease Control and Prevention (CDC) (for serum vitamins A and C), the International Atomic Energy Authority (IAEA) in Vienna, the Community Bureau of Reference of the Commission of the European Communities (BCR) in Belgium (serum proteins), and the U.K. National Institute of Biological Standards and Controls (serum ferritin, soluble transferrin receptor). A reference material for erythrocyte enzymes for vitamin B_6, riboflavin, and thiamin is also available from the Wolfson Research Laboratory, Birmingham, England.

The importance of the use of SRMs is highlighted by the discrepancies in serum folate and red blood cell folate based on

the radioprotein-binding assay (RPBA) and a microbiological assay. By using the newly available SRM for folate, U.S. NHANES established that values based on the RPBA assay were 25–40% lower for serum folate and 45% lower for red blood cell folate compared to both the microbiological method and that using liquid chromatography-tandem mass spectrometry. Because most of the cutoffs to assess the adequacy of folate status were established using the RPBA assay, applying such mismatched cutoffs for the microbiological assay resulted in risks of folate deficiency which were markedly higher (i.e., 16% vs 5.6% for serum folate and 28% vs. 7.4% for RBC folate) (Pfeiffer et al., 2016). These data emphasize the importance of using accurate analytical methods and applying method-specific cutoffs to avoid misinterpretation of the data (MacFarlane, 2016).

If suitable reference materials are not available, aliquots from a single homogeneous pooled test sample should be analyzed by several independent laboratories using different methods. Programs are available which compare the performance of different laboratories in relation to specific analytical methods. Some examples include the programs operated by IAEA, the Toxicology Centre in Québec, Canada, and the U.S. National Institute of Standards and Technology (NIST).

Important differences distinguish assays undertaken by a hospital clinical laboratory from those completed during a survey or research study. Clinical laboratories often focus on values for the assay that are *outside* the normal range, whereas in nutrition surveys such as U.S. NHANES III), and in research studies, the emphasis is often on concentrations that fall *within* the normal range. This latter emphasis requires an even more rigorous level of internal laboratory quality control (Potischman, 2003).

Box 15.10 U.S. NHANES III laboratory quality-control procedures

- Bench quality-control pools for each analyte, at multiple concentration levels
- Blind quality-control pools for each analyte, low-normal and high-normal levels
- Random re-analysis of 5% of specimens for each method
- Split-duplicates from one original specimen submitted from the mobile examination center
- Re-collection from sample participants to provide two observations for comparison of values
- External proficiency testing for many analytes, such as the College of American Pathologists, New York State, CDC-Wisconsin programs.

Where possible, it is preferable for all specimens to be analyzed in a single batch to reduce between-assay variability. This is not always feasible: in such cases, an appropriate number of controls should be included in each batch of samples. Box 15.10 highlights the procedures adopted in U.S. NHANES III to ensure analytical accuracy (Gunter and McQuillan, 1990).

Most clinical chemistry laboratories are required to belong to a certified quality assurance program. The U.S. CDC operates a National Public Health Performance Standards Program (NPHPSP), designed to improve the quality of public health practice and performance of public health systems, particularly statewide assessments.

15.5.8 Predictive value

The predictive value describes the ability of a nutritional biomarker, when used with an associated cutoff, to predict correctly the presence or absence of a nutrient deficiency or disease. Numerically, the predictive value of a biomarker is the proportion of all results of the biomarkers that are true (i.e., the sum of the true positives and true negatives divided by the total number of tests). Because it incorporates information on both the biomarker and the population being tested, predictive value is a good measure of overall clinical usefulness.

The predictive value can be further subdivided into the positive predictive value and the negative predictive value. The positive predictive value of a biomarker is the proportion of positive biomarker results that are true (the true positives divided by the sum of the true positives and false positives). The negative predictive value of a biomarker is the proportion of negative biomarker results that are true (the true negatives divided by the sum of the true negatives and false negatives). In other words, the positive predictive value is the probability of a deficiency state in an individual with an abnormal result, whereas the negative predictive value is the probability of an individual not having the condition when the biomarker result is negative.

Sensitivity, specificity, and prevalence of the nutrient deficiency or disease affect the predictive value of a biomarker. Of the three, prevalence has the most influence on the predictive value of a biomarker. When the prevalence of the condition is low, even very sensitive and specific biomarker tests have a relatively low positive predictive value. In general, the highest predictive value is achieved when specificity is high, irrespective of sensitivity.

15.5.9 Scoring criteria to select biomarkers

European researchers have developed a set of criteria which can be used to select the appropriate biomarkers in nutrition research (Calder et al., 2017), and these criteria are shown in Box 15.11. Once the biomarker has been assessed by applying these criteria, then the information obtained can be used to score the biomarker to determine its usefulness. Details of the proposed scoring system are given in Calder et al. (2017).

Box 15.11. Scoring criteria to select biomarkers

- **Methodological aspects, excluding study design**
  - Method should be validated according to recognized guidelines
  - Appropriate sensitivity
  - Appropriate specificity
  - Reproducibility, accuracy, standardization, stability (quality of sample) and technical variation
  - Biological variation
- **Reflects/marks the biological purpose of the biomarker**
  - A change in the biomarker is linked with a change in the endpoint in one or more target populations
  - Method should be validated according to recognized guidelines
- **Relevance to nutrition research**
  - What is considered as a normal range for healthy people?
  - What is a significant change (consider both biological and statistical)?
  - Is there evidence that nutrition influences the marker? If so, what is the size of the effect reported?
  - Which other factors also have an effect on the biomarker (if any)
  - Are there experimental data where dietary intervention has not resulted in an anticipated change?

Modified from Calder et al. (2017).

In general, a combination of biomarkers should be used where possible rather than a single biomarker for each nutrient; several concordant abnormal values are more reliable than a single aberrant value in diagnosing a deficiency state. Table 15.7 summarizes the biomarkers recommended by the BOND Expert Panels for the assessment of the six micronutrients (folate, iodine, iron, vitamin A, vitamin B_{12}, zinc) of public health importance in low- and middle-income countries (LMICs).

Table 15.7. The recommended biomarkers for six micronutrients of public health importance. RBC: red blood cell; DBS: dried blood s table is drawn from six "Biomarkers of Nutrition for Development Reviews" (Folate, Iodine, Iron, Vitamin A, Vitamin B_{12}, Zinc).

Nut-rient	Biomarkers of Exposure	Biomarkers of status	Functional bio
Folate	Dietary folate equivalents	Serum folate; RBC folate: MBA method	Plasma homoc
Iodine	Salt iodine	Urinary iodine	Thyroglobulin
Iron	Bioavailable iron intakes	Ferritin; RBC proto-porphyrin; transferrin receptor; body Iron index	Currently no bi deficiency
Vitamin A	Dietary vitamin A as retinol activity equivalents (RAE)	Retinol in plasma, DBS, & breast milk; Retinol binding protein in plasma or DBS	Modified relati adaptation Pup
Vitamin B_{12}	Dietary B_{12} intake	Serum B_{12}; Serum holoTC	Serum methyl-l homo-cysteine
Zinc	Dietary Zn intakes; Absorbable Zn	Serum zinc	Impaired linea

This page titled 13.5: Factors affecting choice of nutritional biomarkers (15.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13.6: Evaluation of the selected nutritional biomarker (15.6)

At the population level, nutritional biomarkers are often used for surveys, screening, surveillance, monitoring, and evaluation (Box 15.1), when they are evaluated by comparison with a distribution of reference values from a reference sample group (if available) using percentiles or standard deviation scores.

Alternatively, individuals in the population can be classified as "at risk" by comparing biomarker values with either statistically predetermined reference limits drawn from the reference distribution, or with clinically or functionally defined "cutoff points". At the population level, the biomarkers do not necessarily provide certainty with regard to the status of every individual in the population. In contrast, when using biomarkers for the diagnosis, treatment, follow-up, or counseling of individual patients, their evaluation needs to be more precise, with cut-offs chosen accordingly(Raghaven et al., 2016). Note that statistically defined "reference limits" are technically not the same as clinically or functionally defined "cutoffs", and the two terms should not be used interchangeably.

15.6.1 Reference distribution

Table 15.8. Selected percentiles for hemoglobin (g/dL) and transferrin saturation (%) for male subjects (all races)20–64y. Percentiles are for the U.S. NHANES II "reference population". Abstracted from Pilch and Senti, 1984.

Males	**hemoglobin percentiles (g/dL)**						
Age (y)	5	10	25	50	75	90	95
20–44	13.7	14.0	14.6	15.3	15.9	16.5	16.8
45–64	13.5	13.8	14.4	15.1	15.8	16.4	16.8
Males	**Transferrin saturation percentiles (%)**						
Age (y)	5	10	25	50	75	90	95
20–44	16.6	18.4	23.3	29.1	35.9	43.7	48.5
45–64	15.2	17.6	21.8	27.8	34.2	39.7	44.4

Normally, evaluation at the population level requires a distribution of reference values obtained from a cross-sectional analysis of a reference sample group. Theoretically, only healthy persons free from conditions known to affect the status of the nutrient under study are included in the reference sample group. For example, a distribution of reference values for hemoglobin (by age, sex, and race) was compiled from the U.S. NHANES II based on a sample of healthy, non-pregnant individuals. Participants with conditions known to affect iron status, such as pregnant women and those who had been pregnant in the preceding year, those with white blood cell count < 3.4×10^9/L or > 11.5×10^9/L, with protoporphyrin > 70µg/dL red blood cells, with transferrin saturation < 16%, or with a mean corpuscular volume < 80.0 or > 96.0fL, were excluded Pilch and Senti, 1984. Table 15.8 shows the percentiles for hemoglobin and transferrin saturation for male subjects (all races) aged 20–64y, drawn from the U.S. NHANES II healthy reference sample Pilch and Senti, 1984. A more detailed discussion of the selection criteria for a reference sample group can be found in Ichihara et al. (2017). These distributions can be used as a standard for comparison with the hemoglobin distributions from other study population surveys.

As an example, if anemia is present in the study population, the hemoglobin distribution will be shifted to the left, as shown in the school-aged children from Zanzibar when compared to the optimal hemoglobin distribution for the U.S. NHANES II reference population of healthy African American children in Figure 15.13.

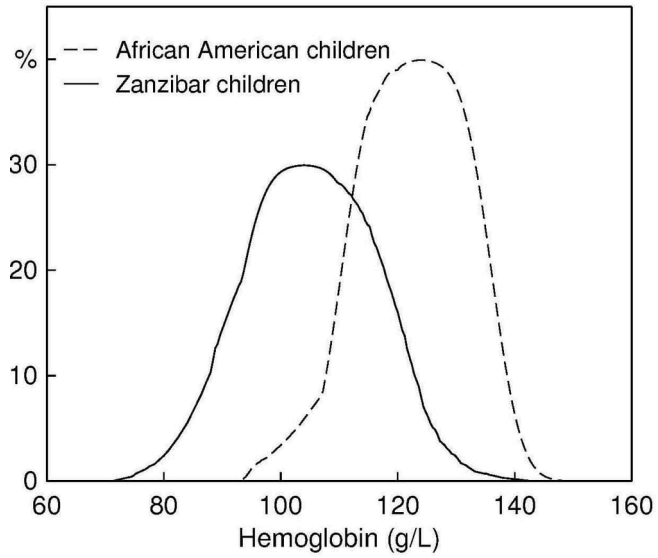


Figure 15.13. Distribution of hemoglobin in a reference population of African American children with no biochemical signs of iron deficiency and school-age children from Zanzibar. Redrawn from Yip et al., 1996.

Distributions of serum zinc concentrations from the U.S. NHANES II survey based on a healthy reference sample (Pilch and Senti, 1984) were developed by Hotz et al. (2003). Data for individuals with conditions known to significantly affect serum zinc concentrations were excluded,i. e., those with low serum albumin (< 35g/L), those with an elevated white blood cell count (> 11.5×10^9/L), and those using oral contraceptive agents, hormones or steroids, or experiencing diarrhea. The International Zinc Consultative Group (IZiNCG) also took age, sex, fasting status (i.e., > 8h since last meal), and time of day of the blood sample collection into account, in the reanalysis. From these data, distributions of reference values for serum zinc (by age, sex, fasting status and time of sampling) were compiled.

Unfortunately, none of the other biochemical data generated from U.S. NHANES II or U.S. NHANES III have been treated in this way(Looker et al., 1997). As a consequence, and in practice, the reference sample group used to derive the values for the reference distribution is usually drawn from the "apparently healthy" general population sampled during nationally representative surveys and assumed to be disease-free. For example, Ganji and Kafai (2006) compiled population reference values in this way for plasma homocysteine concentrations for U.S. adults by sex and age in non-Hispanic whites, in non-Hispanic blacks, in Mexican Americans and in Hispanic subjects using data from U.S. NHANES 1999–2001 and 2001–2002.

15.6.2 Reference limits

The reference distribution can also be used to statistically derive reference limits and also to derive a reference interval. Two reference limits are often defined, and the interval between and including them is termed the "reference interval". On average, 120 "healthy" individuals are needed to generate the reference limits for subgroups within strata such as age group, sex, and possibly race (Lahti et al., 2002). The reference interval usually includes the central 95% of reference values, and is often termed the "reference range" or "range of normal", with the lower 2.5^{th} percentile value often corresponding to the lower reference limit and the upper 2.5^{th} percentile value to the upper reference limit. For example, the reference limit determined by IZiNCG was based on the 2.5^{th} percentile of serum zinc concentrations for males and females aged < 10y and ≥ 10y, qualified by fasting status and time of blood collection (Hotz et al., 2003). Similarly, in the U.K. national surveys, the lower reference limit for hemoglobin was represented by the 2.5^{th} percentile qualified by age and sex. The number and percentage of individuals with observed values falling below the 2.5^{th} percentile value can then be calculated.

Box 15.12 depicts the relationship between reference values, the reference distribution, and reference limits, and how reference samples are used to compile these values. Observed values for individuals in the survey are classified as "unusually low", "usual", or "unusually high", according to whether they are situated below the lower reference limit, between or equal to either of the

reference limits, or above the upper reference limit.

Box 15.12. The relationship between the reference population, the reference distribution, and reference limits

- REFERENCE INDIVIDUALS
  ↓ make up a
- REFERENCE population
  ↓ from which is selected a
- REFERENCE SAMPLE GROUP
  ↓ on which are determined
- REFERENCE VALUES
  ↓ on which is observed a
- REFERENCE DISTRIBUTION
  ↓ from which are calculated
- REFERENCE LIMITS
  ↓ that may define
- REFERENCE INTERVALS

From IFCC (1987).

Unfortunately, no data are available from national nutrition surveys for the distribution of reference values for most functional physiological biomarkers (e.g., relative dose-response for vitamin A) with the exception of child growth (de Onis et al, 2008), and six child gross motor milestones (MGRS, 2004). and for behavioral biomarkers (e.g., cognition; depression). Use of such biomarkers (with the exception of growth) is generally not feasible in large-scale nutrition surveys. Consequently, these functional biomarkers are often evaluated by monitoring their improvement serially, during a nutrition intervention program. Alternatively, observational studies have examined correlations between a static or functional biomarker of a nutrient and a physiological or behavioral biomarker. These observational studies have comprised cross-sectional, case-control, and cohort studies.

The observed values may also be compared using cutoff points as described below.

15.6.3 Cutoff points

Cutoff points, unlike statistically defined reference limits, are based on the relationship between a nutritional biomarker and low body stores, functional impairment or clinical signs of deficiency or excess (Raghavan et al., 2016). The Institute of Medicine (IOM) defines a cutoff for a biomarker as a "specified quantitative measure used to demarcate the presence or absence of a health-related condition often used in interpreting measures obtained from analyses of blood" (IOM, 2010).

The use of cut-off points is less frequent than that of reference limits because information relating biomarkers and functional impairment or clinical signs of deficiency or excess is often not available. Cutoff points may vary with the local setting because relationships between the biomarkers and functional outcomes is unlikely to be the same from area to area.

Cutoff points, like reference limits, are often age-, race-, or sex-specific, depending on the biomarker. For biomarkers based on biochemical tests, cutoff points must also take into account the precision of the assay. Poor precision leads an overlap between those individuals classified as having low or deficient values with those having normal values and thus to misclassification of individuals. This affects the sensitivity and specificity of the test. The International Vitamin A Consultative Group (IVACG), for example, now recommends the use of HPLC for measuring serum retinol concentrations because this is the best method for detecting concentrations < 0.70µmol/L with adequate precision (Tanumihardjo et al., 2016).

The BOND Expert Panel has recommended cutoffs for the biomarkers of exposure, status, or function for six micronutrients — folate, iodine, iron, vitamin A, vitamin B_{12} and zinc, although for some (e.g., serum zinc), the so-called cutoffs are in fact statistically defined and hence are actually reference limits. Note that in some cases, the so-called cutoffs for status or functional biochemical biomarkers are assay-specific, as discussed earlier for folate. Assay-specific cutoffs are also available for soluble transferrin receptor, a useful biomarker for identifying iron deficiency because it is less strongly affected by inflammation. Assay-specific cutoffs arise when there is no CRM available for the biomarker, as has been the issue for folate and soluble transferrin receptor, until recently. Different cutoff units are sometimes used, presenting an additional challenge when interpreting data

across laboratories (Raghavan et al., 2016).

Table 15.9 highlights how the use of differing cutoffs could affect the prevalence of vitamin B_{12}deficiency in a sample of U.S. elderly people (at least 60y) participating in the U.S. NHANES surveys, with estimates ranging from 3% to 26% (Yetley et al., 2011). This means that in population studies, prevalence estimates for deficiency can vary according to the cutoff applied, which has implications for nutrition public policy, and which may result in unwarranted clinical interventions. In some cases, cutoff points for a nutritional biomarker vary according to the functional outcome of interest. In the elderly, for example, maximum muscle function has been associated with 25-OH-D levels > 65nmol/L, whereas reduction in fracture risk is associated with higher serum 25-OH-D levels (Dawson-Hughes et al., 2008).

Table 15.9. Prevalence estimates of vitamin B_{12} deficiency using different cutoffs for serum vitamin B_{12}. From Raghavan et al.(2016).

Cutoffs. Serum vitamin B_{12} pmol/L	Prevalence estimates for for vitamin B_{12} def
< 148	2.9 ± 0.2
< 200	10.6 ± 0.4
< 258	25.7 ± 0.6

Receiver operating characteristic (ROC) curves are used to evaluate the ability of a nutritional biomarker to classify individuals with the condition under investigation. The curves portray graphically the trade-offs that occur in the sensitivity and specificity of a biomarker when the cutoffs are altered. To use this approach, a spectrum of cutoffs over the observed range of biomarker results is used, and the sensitivity and specificity for each cutoff calculated. Next, the sensitivity (or true-positive) rate is plotted on the vertical axis against the true negative rate (1-specificity) on the horizontal axis for each cutoff, as shown in Figure 15.14, The closer the curve follows the left-hand border and then the top-border of the ROC space, the more accurate is the biomarker cutoff in distinguishing a deficiency from optimal status. The optimal ROC curve is the line connecting the points highest and farthest to the left of the upper corner. The closer the curve comes to the 45° diagonal of the ROC space, the less accurate the biomarker cutoff (Søreide, 2009). Most statistical programs (e.g., SPSS) provide ROC curve analysis.

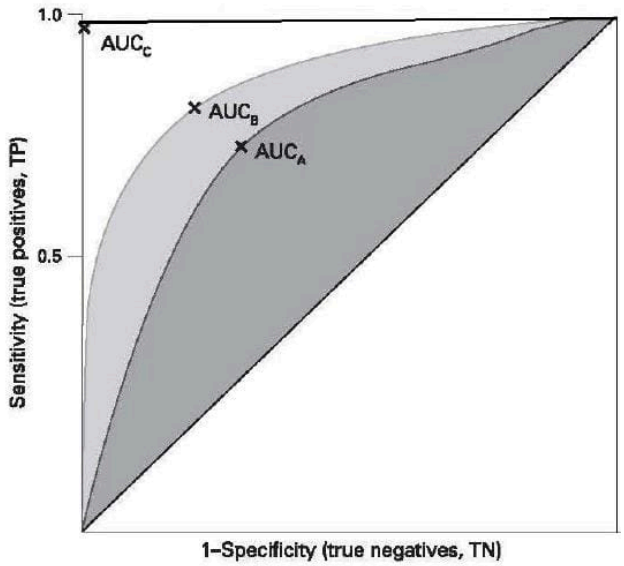


Figure 15.14. Receiver-operating characteristic curves. Three plots and their respective areas under the curve (AUC) are given. The diagnostic accuracy of marker C (white area) is better than that of B and A, as the AUC of C > B > A. X = point of best cut-off for the biomarker. From: Søreide, 2009, with permission of the BMJ Publishing Group Ltd.

The area under the ROC curve (AUC), also known as the cut-point "*c*" statistic or c-index, is a commonly used summary measure of the accuracy of the biomarker cutoff. AUCs can range from 0.5 (random chance, or no predictive ability — a 45° line on the ROC plot, Figure 15.4) to 0.75 (good), and to > 0.9 (excellent). The cutoff value that provides the highest sensitivity and specificity is calculated. On the rare occasions that the estimated AUC for the biomarker cutoff is < 0.5, then the biomarker cutoff is worse than chance! When multiple biomarkers are available for the same nutrient, the biomarker with the highest AUC is often selected.

Youden index (J) is another summary statistic of the ROC curve used in the interpretation and evaluation of biomarkers. It defines the maximum potential effectiveness of a biomarker. The statistic J can be defined as J = (maximum sensitivity (c) + specificity (c) – 1). The cut-off that achieves the maximum is referred to as the optimal cutoff (c*) because it is the cut-off that optimizes the biomarker's differentiating ability when equal weight is given to sensitivity and specificity. The statistic J can range from 0 to 1, with 1 indicating a perfect diagnostic test, whereas values closer to 0 signify a limited effectiveness (Schisterman et al., 2005; Ruopp et al., 2008).

Misclassification arises when there is overlap between individuals who actually have the deficiency and those falsely identified (i.e., false positives). Neither reference limits nor cutoff values can separate the "deficient" and the "adequately nourished" without some misclassification occurring. This is shown in Figure 15.15 for the real-life situation (B). Note that the cut-offs finally selected can vary according to whether the consequences of a high number of individuals being falsely classified as positive is more or less important than the consequences of a large number of individuals being falsely classified as negatives. Minimizing either misclassification may be considered more important than minimizing the total number of individuals misclassified.

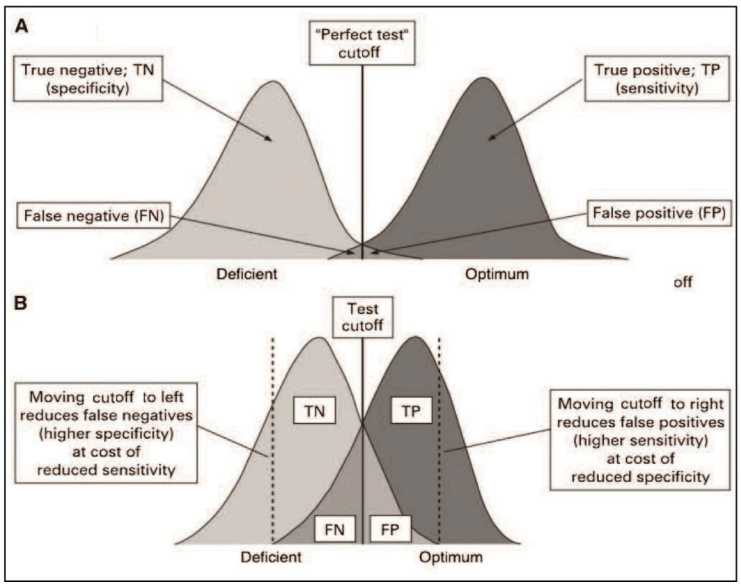


Figure 15.15. A good discriminatory test with almost perfect ability to discriminate between people with a nutrient deficiency and those with optimum nutrient status. The ability to correctly detect all the true negatives depends on the specificity of the biomarker; the ability to correctly detect all the true positives depends on the sensitivity of the biomarker. FN, false negative; FP, false positive; TN, true negative; TP,true positive. From: Raghaven et al.(2016), Reproduced by permission of Oxford University Press on behalf of the American Society for Nutrition.

Note that the sensitivity can be improved (i.e., reducing the false positives) by moving the cut-off to the right but this reduces the specificity (false negatives), whereas moving the cut-off to the left reduces the false negatives (higher specificity) at the cost of a reduction in sensitivity. The former scenario may be preferred for the clinical diagnosis of a fatal condition, whereas cut-offs with a high specificity may be preferred for diagnostic tests that are invasive or expensive.

Misclassification arises because there is always biological variation among individuals (and hence in the physiological normal levels defined by the biomarker), depending on their nutrient requirements. As well, for many biomarkers there is high within-individual variance, which influences both the sensitivity and specificity of the biomarker, as well as the population prevalence estimates. These estimates can be more accurately determined if the effect of within-individual variation is taken into account. This can only be done by obtaining repeated measurements of the biomarker for each individual on at least a sub-sample of the individuals. The number of repeated measurements required depends on the ratio of the within-individual to between-individual variation for the biomarker and population concerned (see analogous discussion of adjustments to prevalence estimates for inadequate dietary intakes in Chapter 3).

The specificity of the diagnosis can be enhanced by combining biomarkers. The presence of two or more abnormal values can be taken as indicative of deficiency, often improving the specificity of the diagnosis. This approach has been used in several national nutrition surveys for diagnosing iron deficiency, including the U.S. NHANES in 2003. Here a multivariable approach for estimating total-body iron stores was developed based on the ratio of soluble transferrin receptor to serum ferritin (Gupta et al.,

2017). Increasingly, a combined indicator is being used for diagnosing B_{12} deficiency that is based initially on 4 status biomarkers (serum B_{12}, methylmalonic acid (MMA), holotranscobalamin (holoTC), and total homocysteine (tHcy). However, the indicator can be adapted for use with three or two biomarkers; for more details see Allen et al. (2018).

In the future, a more flexible cutoff approach may be adopted in which two cutoffs are provided, separated by a gray zone. The first cutoff in this gray zone approach is selected to include deficiency with near certainty, while the second is chosen to exclude deficiency with near certainty. When a biomarker falls within the gray zone (suggesting subclinical deficiency), investigators are prompted to seek additional assessment tools in an effort to provide a more precise diagnosis. In this way, unwarranted clinical interventions are avoided (Raghaven et al., 2016).

15.6.4 Trigger levels for surveillance and public health decision making

In population studies, cutoff points may be combined with trigger levels to set the level of an indicator (or a combination of indicators) at which a public health problem exists of a specified level of concern. Trigger levels may highlight regions, populations or sub-groups where specific nutrient deficiencies are likely to occur, or may serve to monitor and evaluate intervention programs. They should, however, be interpreted with caution because they have not always been validated in population-based surveys. Box 15.13 presents examples of trigger levels for zinc biomarkers set by the International Zinc Nutrition Consultative Group (IZiNCG).

Box 15.13. Trigger levels for zinc biomarkers set by the international Zinc Nutrition Consultative Group

- Prevalence of serum zinc less than age/sex/time-of-day specific cutoffs is > 20%
- Prevalence of inadequate zinc intakes below the appropriate estimated average requirements is > 25%
- Prevalence of low height-for-age or length-for-age Z-scores (i.e., < –2SD) is at least 20%.

Note: Ideally, all three indicators should be used together to obtain the best estimate of the risk of zinc deficiency in a population, and to identify specific sub-groups with elevated risk (de Benoist et al., 2007).

WHO (2011) have classified the public health significance of anemia at the population level based on the prevalence of low hemoglobin concentrations. Moreover, reductions in the prevalence of anemia are targets for public health efforts in many low-income countries. To be successful, however, such efforts must address the multifactorial etiology of anemia, and avoid the presumption that anemia is synonymous with nutritional iron deficiency (Raiten et al., 2012).

WHO (2011) defines vitamin A deficiency as a severe public health problem requiring intervention when 20% of children aged 6–71mo have a serum retinol concentration < 0.7µmol/L and another biological indicator of poor vitamin A status. These may include night blindness, breast milk retinol, relative dose response, or modified dose response; or when at least four demographic and ecological risk factors are met; see Tanumiharjo et al.(2016) and Chapter 18 for more details.

Trigger levels to define the severity of iodine deficiency in a population based on total goiter rate have also been defined by WHO (Rohner et al., 2015). The criteria used are < 5%, iodine sufficiency; 5.0–19.9%, mild deficiency; 20–29.9%, moderate deficiency; and > 30%, severe deficiency. Details of the classification system used to diagnose goiter are available in WHO (2007).

A generalized discussion of the specific procedures used for the evaluation of dietary, anthropometric, laboratory, and clinical methods of nutritional assessment are discussed more fully in Chapters 8b, 13, 25, and 26, respectively.

This page titled 13.6: Evaluation of the selected nutritional biomarker (15.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

13.7: Application of new technologies (15.7)

With the development of new technologies, the focus is changing from the use of biomarkers that are associated with specific biochemical pathways to methods that assess the activity of multiple macro- and micro-nutrients and their interactions within complex physiological systems. These new technologies apply “omics” techniques that allow the simultaneous large-scale measurements of multiple genes, proteins, or metabolites, coupled with statistical and bioinformatics tools. Such measurements offer the possibility of characterizing alterations associated with disease conditions, or exposure to food components. However, further work on the development and implementation of appropriate quality control systems for “omics” techniques is required. A brief description of these “omics” techniques and their application in nutritional assessment follows.

15.7.1 Nutrigenetics

Nutrigenetics focuses on understanding how genomic variants interact with dietary factors and the implications of such interactions on health outcomes (Mathers, 2017). Nutrigenetics is being used increasingly to predict the risk of developing chronic diseases, explain their etiology, and personalize nutrition interventions to prevent and treat chronic diseases.

Nutrigenetics uses a combination of recombinant DNA technology, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyse the structure and function of genomes. Genetic variation among individuals is minimal. Nevertheless, there is approximately 1% genetic variation that can lead to a wide variability in health outcomes, depending on dietary intake and other environmental exposures. The most common type of genetic variability among individuals is the single nucleotide polymorphism (SNP), which is a base change in the DNA sequence. With the development of genetic SNP databases, individuals can be screened for genetic variations, some of which can have an effect on an individual's health.

One of the earliest examples is the effect of the common SNP-C677T (A22V) associated with the MTHFR gene. This C677T polymorphism is responsible for a genetic defect in the enzyme methylenetetrahydrofolate reductase (MTHFR) that can cause a severe or a more moderate accumulation of homocysteine. Several studies in both younger and older subjects have shown that individuals homozygous for the MTHFR polymorphism C677T (A222V) have increased levels of plasma homocysteine concentrations, although only in the face of low folate status. No association has been found in homozygotes with adequate folate status. In view of the influence of MTHFR C677T (A222V) polymorphism on plasma homocysteine, this C677T polymorphism has been proposed as an independent risk factor for coronary heart disease (Gibney and Gibney, 2004). Since this early example, there have been several other reports in which polymorphisms have been associated with common chronic diseases through interactions with the intake of both micronutrients and macronutrients, as well as with the consumption of particular foods and dietary patterns.

Chronic diseases such as obesity, type 2 diabetes, and coronary heart disease are probably associated with multiple genetic variants that interact with diet and other environmental exposures. Therefore, predictive testing based on a single genetic marker for these chronic diseases is likely to be of limited value. As a result, increasingly, studies are combining genetic polymorphisms to yield genetic-predisposition scores, often termed genetic risk scores (GRS), in an effort to examine the cumulative effect of SNPs on diet interactions and susceptibility to diseases such as obesity and type 2 diabetes.

As an example, the use of a GRS has been applied in studies examining the interactions between genetic predisposition and consumption of certain foods in relation to body mass index and obesity. In several prospective cohort studies, an interaction between the consumption of sugar-sweetened beverages (Qi et al., 2012), and a GRS based on 32 BMI-associated variants on BMI, has been reported. (Qi et al., 2014). These findings have highlighted the importance of reducing consumption of these foods in individuals genetically predisposed to obesity.

Interactions between dietary patterns and GRS may also be associated with adiposity-related outcomes. In a large study based on 18 cohorts of European ancestry, nominally significant associations were observed between diet score and a GRS based on 14 variants commonly associated with BMI-adjusted waist-hip ratio. Moreover, stronger genetic effects were observed in those individuals with a higher diet score (i.e., those consuming healthier diets) (Nettleton et al., 2015). The clinical relevance of these findings, however, is uncertain, and further experimental and functional studies are required.

Several studies have also examined the effects of GRS on the differential responses to nutrition interventions. Huang et al. (2016), for example, showed that individuals with lower GRS for type 2 diabetes mellitus had greater improvements in insulin resistance

and β-cell function when consuming a low-protein diet. In contrast, individuals with higher GRS for glucose disorders had greater increases in fasting glucose when consuming a high-fat diet (Wang et al., 2016). For more examples of interactions between dietary intakes and genes involved in risk of disease, see Ramos-Lopez et al. (2017).

Clearly, advances in nutrigenetics have the potential to enhance the prediction for risk of developing chronic diseases, as well as personalizing their prevention and treatment. Indeed, increasingly, genetic tests are being used to customize diets based on the predisposition to weight gain by saturated fat intake and the increased risk of developing hypertension by high salt intake.

15.7.2 Proteomics

Proteomics refers to the systematic identification and quantification of the overall protein content of a cell, tissue, or an organism. The proteome is defined as a dynamic collection of proteins that demonstrate variation between individuals, between cell types, and between entities of the same type but under different pathological or physiological conditions (Huber, 2003).

Comparison of proteome profiles between differing physiologic and disease states is used to identify potential biomarkers for the early diagnosis and prognosis of disease states, monitoring disease development, understanding pathogenic mechanisms, and for developing targets for treatment and therapeutic intervention.

Three major steps are involved in proteomics analysis: (i) sample preparation; (ii) separation and purification of complex proteins, and (iii) protein identification. Several methods can be used to separate and purify the samples, including chromatography-based techniques, enzyme-linked immunosorbent assays (ELISA) or Western blotting. More advanced techniques are also being used such as protein microarrays and two-dimensional difference in-gel electrophoresis (DIGE). To identify proteins in great depth, mass-spectrometry-based proteomics is used to measure the highly accurate mass and fragmentation spectra of peptides derived from sequence-specific digestion of proteins. Finally, the raw data from mass spectrometry (MS) are searched using database search engines and software such as MASCOT or Protein-Pilot, etc. For more details of these techniques, see Aslam et al. (2017).

Further work is required to improve the reproducibility and performance of proteomics tools. Systematic errors can be introduced during each step that may artificially discriminate disease from non-disease. Sources of biological and analytical variation have not always been controlled and the sample size for testing a candidate biomarker has sometimes been inadequate. However, with improvements, proteomics has the potential to screen large cohorts for multiple biomarkers, and to identify protein patterns characteristic of particular health or disease states.

15.7.3 Metabolomics

Metabolomics characterizes the small molecular weight molecules, called metabolites, that are present in human biofluids, cells, and tissues at any given time (Brennan, 2013). The aim of metabolomics is to provide an overview of the metabolic status and global biochemical events associated with a cellular or biological system under different biological conditions. The metabolome is comprised of small intermediary molecules and products of metabolism, including those associated with energy storage and utilization, precursors to proteins and carbohydrates, regulators of gene expression, and signalling molecules.

Five major steps are involved in metabolomics: (i) experimental design; (ii) sample preparation; (iii) data acquisition by nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry-based analysis; (iv) data processing; and (v) statistical analyses (O'Gorman and Brennan, 2017). Computational tools have been developed to relate the structure of the metabolites identified to biochemical pathways. This is a complex task as a metabolite may belong to more than one pathway; see Misra (2018).

The biofluids most widely used for metabolomics are blood, urine, and saliva. Several analytical techniques are used to analyze metabolites in these biofluids; each technique has advantages and disadvantages. The major analytical techniques are NMR spectroscopy or mass spectrometry-(MS)-based methods (e.g., gas chromatography (GC)-MS, liquid chromatography (LC)-MS, capillary electrophoresis (CE)-MS) and high performance liquid chromatography (HPLC). No single technique is capable of measuring the entire metabolome.

Both non-targeted and targeted metabolomics can be used, depending on the research question. The non-targeted approach aims to measure as many metabolites as possible in a biological sample simultaneously, thus providing a broad coverage of metabolites, and an opportunity for novel target discovery. In contrast, the targeted approach involves measuring one metabolite

or a specific class of known metabolites with similar chemical structures. This requires the metabolites of interest to be known *a priori* and commercially available in a purified form for use as internal standards so that the amount of a targeted metabolite can be quantified (O'Gorman and Brennan, 2017). Currently, there are no standardized protocols for sample collection and storage for metabolomic studies.

There are three main applications of metabolomics in nutrition research: (i) dietary intervention studies; (ii) diet-related disease studies; and (iii) dietary biomarker studies designed to identify and validate novel biomarkers of nutrient exposure (Brennan, 2013).

Dietary intervention studies can be used to investigate the mechanistic effects of the intervention and to determine the impact of specific foods or diets on metabolic pathways. An example includes the application of metabolomics to investigate the impact of consuming either wholegrain rye bread or refined wheat bread. Metabolomics of serum samples from 33 postmenopausal women indicated that consumption of rye bread decreased the branched chain amino acids leucine and iso-leucine and increased NN-dimethylglycine. Such alterations suggest that wholegrain rye bread may confer beneficial health effects (Moazzami et al., 2012). Consumption of dark chocolate has also been investigated in dietary intervention studies involving metabolomics. In a study by Martin et al. (2009), 30 participants were classified into low and high anxiety traits using validated psychological questionnaires. Participants then received 40g dark chocolate daily for 14d during which urine and plasma were collected at baseline, mid-line, and endline. Consumption of dark chocolate for 14d was reported to reduce stress related molecules in the urine (i.e., cortisol and catecholamines) and partially normalized levels of glycine, citrate, trans-aconitate, proline, and β-alanine in those participants with a high anxiety trait compared to those with a low anxiety trait. These findings indicated alterations in stress-related energy metabolism (Martin et al., 2009).

Diet-related diseases such as type 2 diabetes and cardiovascular disease have been investigated by metabolomics in an effort to understand their etiology and identify new biomarkers. There is now strong evidence that elevated plasma levels of branched chain amino acids (BCAAs) (i.e., leucine, isoleucine, and valine) and their derivatives are linked to the risk of developing insulin resistance and type 2 diabetes. Moreover, depending on the metabolite, changes may be apparent as long as 13 years ahead of clinical manifestations of type 2 diabetes. Several investigators have also shown that BCAAs and related metabolites are also associated with coronary heart disease, even when controlled for diabetes. See Newgard (2012), Klein and Shearer (2016), and Bhattacharya et al. (2014) for more details.

With the accumulating evidence of the importance of the gut microbiota in the development of certain diseases, metabolomics is also being used to identify metabolites that originate from gut microbial metabolism, and follow alterations that may occur (Brennan, 2013).

Nevertheless, some of the findings reported from metabolomics studies have been contradictory, highlighting the need for further research before metabolomic results can be translated into clinical applications.

Dietary biomarker studies are being explored to overcome some of the limitations of traditional dietary assessment methods, and thus improve the assessment of the relationship between diet and chronic disease. The food metabolome (i.e., metabolites derived from foods, and food constituents) is a promising resource to discover novel food biomarkers.

Several approaches are used to identify novel biomarkers of dietary intake. They may involve acute feeding studies and short- to medium-term dietary intervention studies in a controlled setting. These intervention studies focus only on one or a few specific type(s) of food(s), after which biofluids, most notably urine or serum, are collected postprandially or following the short-to medium term dietary intervention. This approach has been used to identify several putative biomarkers of specific foods and drinks such as citrus fruits, cruciferous vegetables, red meat, coffee, sugar-sweetened beverages, and wine (O'Gorman and Brennan, 2017). However, the biomarkers identified in this way reveal no information on other dietary origins of the identified biomarkers. In addition, for those biomarkers that are short-term, and excreted rapidly and almost completely over 24hr, their usefulness as biomarkers of habitual intake remains questionable.

Potential biomarkers for specific foods can also be identified using cohort studies. In this approach, metabolic profiles of high or low consumers of specific food(s), identified by a self-reported dietary questionnaire are examined. Studies of this type have identified proline betaine and flavanone glucuronides as potential biomarkers of citrus fruit intake (Pujos-Guillot et al., 2013). Biomarkers for fish, red meat, whole-grain bread, and walnuts have also been identified using this approach (O'Gorman and Brennan, 2017). Nevertheless, because these studies only generate associations, validation of the metabolite as a specific biomarker of intake should be confirmed through a controlled dietary intervention study.

Large cross-sectional or cohort studies have also used dietary patterns to identify multiple biomarkers of food intake. Dietary

patterns can be identified by principal component analysis or k-means cluster analysis. Once identified, the dietary patterns are linked to metabolomic profiles through regression (or other statistical methods) to identify dietary biomarkers. Using this approach, metabolites have been identified that can be used to predict compliance to complex diets and to study relationships between diet and disease (Bouchard-Mercier et al., 2013). In some of these studies, the predictive accuracy of the identified biomarkers has been evaluated through the use of receiver operating characteristic (ROC) analysis (Heinzmann et al., 2012; Wang et al., 2018).

Once identified, the performance of all biomarkers of dietary exposure must always be validated in an independent and diverse epidemiological study, and across different laboratories to establish whether they are generalizable to free-living populations. This approach was used by Heinzmann et al.(2012) to validate proline betaine as a biomarker of citrus intake. In addition, the suitability of the biomarker over a range of intakes should be confirmed through a dose-response relationship (O'Gorman and Brennan, 2017).

This page titled 13.7: Application of new technologies (15.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14: IRON (CHAPTER 17)

Abstract

Dietary iron is present as heme and nonheme iron, each absorbed by different mechanisms. Once absorbed, the iron released can be stored or bound to plasma transferrin for distribution to other tissues. Quantitatively, most iron is used by immature red blood cells in the bone marrow for hemoglobin (Hb) production. Senescent erythrocytes are degraded by macrophages, and the iron released from catabolized Hb re-enters the circulation. Absorption of iron is modulated in response to the level of body iron stores and by the amount of iron needed for erythropoiesis. Hepatic hepcidin is the master regulator of iron homeostasis. Hepcidin levels are suppressed by iron depletion or increased iron demand (i.e., enhanced erythropoiesis), thereby increasing absorption and mobilization of iron from body stores into plasma. In contrast, when iron stores are replete (and during inflammation), hepcidin levels are increased, so iron cannot efflux into the circulation, thus preventing iron overload.

Anemia is the most common sign of iron deficiency in low- and middle-income countries, with with infants, young children, and women of child-bearing age at greatest risk. Features of iron deficiency anemia (IDA) include impairments in work capacity and cognition, and possibly risk during pregnancy of low birthweight and prematurity. Despite low specificity and sensitivity, a low Hb concentration is the most widely used test for IDA. Development of IDA occurs in three stages, the first being a decrease in iron stores, reflected by a low serum ferritin, followed by iron-deficient erythropoiesis. At this stage, iron stores are exhausted (i.e., serum ferritin < 12µg/L), iron supply to the erythropoietic cells is progressively reduced, and decreases in transferrin saturation occur. At the same time, serum soluble transferrin receptor (sTfR) and erythrocyte protoporphyrin increase. Only in the third stage is there a decline in Hb, decreases in hematocrit and red-cell indices, and frank microcytic, hypochromic anemia, confirmed by examination of a stained blood film.

This chapter describes how to assess the adequacy of dietary iron intakes, followed by details of the hematological parameters used to diagnose anemia, the serum biomarkers to identify iron depletion (serum ferritin), and iron deficient erythropoiesis (serum ferritin; serum iron; transferrin saturation; sTfR; and erythrocyte protoporphyrin). Advantages and limitations are discussed together with details of the measurement and interpretive criteria for each biomarker. Use of a regression modeling approach to adjust for the effect of inflammation on serum ferritin and sTfR is highlighted, in view of the challenge of distinguishing between IDA and anemia of chronic disease. The final section emphasizes the simultaneous use of multiple iron biomarkers to provide a more valid assessment of iron status and minimize misclassification that may occur when using a single measure. The advantages of using the “Total Body Iron Model” based on serum ferritin and sTfR expressed as body iron (mg/kg), with a cutoff of < 0mg/kg to define iron deficiency, is described. Finally, details of emerging iron indicators, notably hepcidin, non-transferrin-bound iron, and some reticulocyte indices, are presented.

14.1: Introduction and functions of iron (17.1)

14.2: Hemoglobin (17.2)

14.3: Hematocrit or packed cell volume

14.4: Red-cell indices

14.5: Red-cell distribution width (17.5)

14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)

14.7: Serum ferritin (17.7)

14.8: Erythrocyte protoporphyrin

14.9: Serum soluble transferrin receptor (17.9)

14.10: Multiple biomarkers (17.10)

14.11: Emerging iron status indicators (17.11)

This page titled 14: Iron (Chapter 17) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.1: Introduction and functions of iron (17.1)

The assessment of the iron status of the population is critical: iron deficiency is a widely recognized micronutrient deficiency, and a public health problem worldwide.

The iron content of the human body is carefully regulated, normally containing about 3 to 4g of iron. Of this, about 65% is present in hemoglobin (Hb), the oxygen-carrying pigment of the red blood cells, that plays a critical role in transferring oxygen from lung to tissues. Hemoglobin is made up of four heme subunits, each with a polypeptide chain of globin attached. Each molecule of heme consists of a protoporphyrin IX molecule with one iron atom. In addition, about 3.5% of body iron is present in myoglobin, the oxygen-binding storage protein found in muscle. The structure of myoglobin is similar to Hb, except that it contains only one heme unit and one globin chain.

17.1.1 Distribution of body iron in men and women

Table 17.1 shows the amount of iron and its distribution in the body in males and females. Trace amounts of iron are also associated with electron transport and several enzymes. Examples include the heme-containing cytochromes that serve as electron carriers within the cell, iron-sulfur proteins (flavoproteins, heme-flavoproteins) that are required for the first reaction in the electron transport chain, and hydrogen peroxidases (e.g., catalase and peroxidase). The cytochrome-P450 family of enzymes also contain heme and are located in microsomal membranes of liver cells and intestinal mucosal cells. Key functions of cytochrome-P450 involve detoxification of foreign substances in the liver, and synthesis of prostaglandins, steroid hormones, and bile acids (Yip and Dallman, 1996; Beard et al., 1996).

Table 17.1. Distribution of body iron in men and women. Data from Elsayed et al.(2016).

Amount of Iron	Male (mg)	Female (mg)	% of To
Total	3000-4000	2000-3000	100
Hemoglobin	2400	1700	65
Intracellular storage (ferritin and hemosiderin)	1000	500	30
Myoglobin	150	120	3.5
Transferrin-bound iron	4	3	0.1

In addition to these functional forms, as much as 30% of total body iron is present as storage iron, found primarily in the liver. Smaller amounts occur in the reticulo-endothelial cells of the bone marrow and spleen; and in the muscle tissues. Of the storage iron, approximately two-thirds consists of ferritin, the soluble fraction of the nonheme iron stores. The remainder of storage iron is insoluble hemosiderin. Small quantities of ferritin can be synthesized in all cells of the body, even those with no special iron storage function. Ferritin also appears in small concentrations in the serum but is not involved in iron transport. In healthy individuals with no evidence of inflammation, serum ferritin concentration is closely correlated with the size of the iron store (Yip and Dallman, 1996).

Stored iron serves as a reservoir to supply cellular needs, mainly Hb production, and is especially important in the third trimester of pregnancy. The size of the storage component is most strongly influenced by age, sex, body size, and either the magnitude of iron losses or the presence of diseases of iron overload (Brittenham et al., 1981).

Iron transport is carried out by the transport protein transferrin, which comprises only about 0.1% of total body iron (Table 17.1). Transferrin delivers iron to the tissues by means of cell membrane receptors specific for transferrin (i.e., sTfR1 and sTfR2). About 20-30mg of iron cycles through the transport component each day.

17.1.2 Absorption and metabolism of iron

Four main factors in the body operate to maintain iron balance and prevent iron deficiency and iron overload. These are: (a) intake,

(b) absorption (c) storage, and (d) loss of iron. The interrelationship of these factors has now been described mathematically, so that the amount of storage iron can be predicted as long as iron losses and bioavailability of iron are known (Hallberg et al., 1998).

In an iron-sufficient adult male, absorption of iron from the diet only contributes about 1mg/d, while for menstruating women, about 1.5mg/d is absorbed to compensate for menstrual blood loss. There are two different forms of iron in the diet, heme and nonheme, each absorbed primarily in the duodenum. Heme iron is derived mainly from Hb and myoglobin in red meat, poultry, and fish and is absorbed readily as the intact iron porphyrin. Absorption of heme iron ranges from 15-35% and is little affected by dietary factors; absorption is regulated primarily in response to total body iron. As demand for body iron increases there is a corresponding upregulation in the uptake of heme iron and also in the rate of transfer to the duodenal enterocytes. Details of the exact mechanism controlling heme iron absorption are poorly understood (Anderson and Frazer, 2017). However, once absorbed, iron ($Fe2^+$) is released from heme iron by heme oxygenase, after which the iron follows the same pathway as nonheme iron, as described below(EFSA, 2015).

Nonheme iron (i.e., inorganic iron) is found primarily in plant-based foods, but also in meat, eggs, and some dairy foods. Liver, iron-fortified foods, and the seeds of legumes, such as beans, contain nonheme ion in the form of ferritin, although the mechanism whereby ferritin iron is absorbed is poorly understood (EFSA, 2015). Plant ferritin is readily released during cooking and digestion (Lynch et al., 2018). The absorption of nonheme iron is less efficient than heme iron and can range from 2-20%, depending on the iron status of the individual and the simultaneous ingestion of other dietary components; some inhibit and others enhance nonheme iron absorption. Most of the food components classed as inhibitors generally act by binding iron in the gastrointestinal tract, preventing its absorption. In contrast, the enhancers act by forming complexes that can be taken up by the intestinal iron transport proteins, thus preventing the iron from binding to the inhibitors. Alternatively, they act by reducing the reactive ferric ($Fe3^+$) iron to its less reactive and more soluble ferrous ($Fe2^+$) state.

Inhibitors of nonheme iron absorption include phytic acid, polyphenols, calcium, and peptides from partially digested proteins (e.g., soybean protein). Of these, phytic acid (myoinositol hexa-phosphate, IP6) is the main inhibitor of nonheme iron absorption from plant-based diets. However, traditional food processing methods such as soaking, germination, and fermentation can lower the phytic acid content, either by reducing the water-soluble phytate content, or by phytase hydrolysis converting the phytic acid to lower myo-inositol phosphate forms that no longer inhibit nonheme iron absorption (i.e., IP2; IP1) (Gibson et al., 2018). Polyphenol compounds from beverages such as tea, coffee, cocoa, and red wine, vegetables (spinach, aubergine), legumes (colored beans), and cereals such as red sorghum, also inhibit nonheme iron absorption. Their inhibitory effect is dose-dependent, the strength depending on the structure of the phenolic compounds, with the gallate-containing tea polyphenols having the largest inhibitory effect (Hurrell et al., 1999). Calcium inhibits both nonheme and heme iron absorption, although the inhibitory effect is weak and short-term, with a mixed effect on iron status and no reduction in Hb (Abioye et al., 2021). The mechanism is not well understood (Lynch et al., 2018).

In contrast, ascorbic acid and muscle tissue from meat, poultry, fish and liver, enhance nonheme iron absorption when eaten in meals containing inhibitors (Lynch et al., 2018). Of these, ascorbic acid is the most potent enhancer of nonheme iron absorption through its ability to convert ferric ($Fe3^+$) iron to ferrous ($Fe2^+$) iron at low pH as well as its chelating properties (Conrad and Scade, 1968). The enhancing effect of ascorbic acid is dose-dependent and most marked when consumed with meals containing high levels of inhibitors, including phytate and polyphenols. Ascorbic acid can also enhance absorption of many iron fortification compounds (Hurrell et al., 2004), except NaFeEDTA (Troesch et al., 2009). The mechanism whereby muscle tissue enhances nonheme iron absorption is not clear. It may be linked to cysteine-containing peptides released during digestion, which have the potential to reduce ferric ($Fe3^+$) iron to ferrous ($Fe2^+$) iron and chelate iron as described for ascorbic acid (Taylor et al., 1986).

The intestinal absorption of nonheme iron has been extensively studied; for details see Anderson and Frazer (2017). Briefly, the nonheme iron in foods that exists predominantly as ferric ($Fe3^+$) iron is solubilized in the gastrointestinal lumen, where first it is reduced by duodenal cytochrome b reductase (DCYTB) to ferrous iron ($Fe2^+$) before being transported by the iron-import protein divalent metal-ion transporter 1 (DMT1) into the enterocytes of the small intestine. When demand for iron is low, the iron can be stored within the cell as the iron storage protein ferritin. When the demand for body iron is high, iron is either taken up by mitochondria for the synthesis of heme, or exported into the circulation by the iron export protein ferroportin 1 (FPN1). To leave the cell and bind to plasma transferrin, however, $Fe2^+$ must be oxidized back to $Fe3^+$. This oxidation is catalyzed by a membrane-bound ferroxidase hephaestin. Transferrin-bound iron is then distributed via the circulation throughout the body to sites of utilization (Figure 17.1). Most of the iron is used by immature red blood cells in the bone marrow for Hb production.

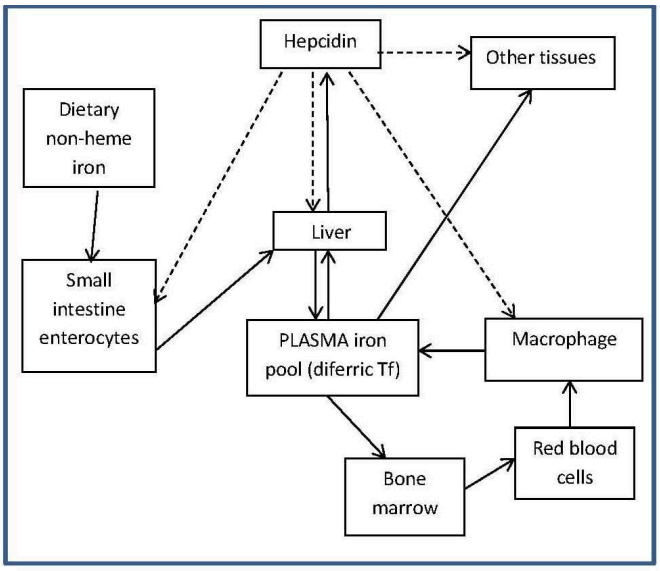


Figure 17.1. Body iron homeostasis. Dietary iron is present as both heme and nonheme iron. The mechanisms underlying heme absorption are poorly understood. Dietary non-heme iron is first reduced to the ferrous form by DCYTB and then it is transported by DMT1 into the small intestinal enterocytes. When demand for iron is low, iron is stored as ferritin, mainly in the liver whereas when demand is high, iron is exported into the plasma by FPN1 where it is transported bound to transferrin, and delivered to the target tissues: bone marrow, liver, and other tissues, such as muscle. Quantitatively, most iron is used by immature red blood cells in the bone marrow for Hb production in red blood cells. Senescent erythrocytes are phagocytosed by macrophages, and the iron is released from catabolized Hb and re-enters the plasma. Hepcidin plays a critical role in regulating whole body iron levels and is regulated by body iron demand.
DCYTB, duodenal cytochrome b;
DMT1, divalent metal-ion transporter 1;
FPN1, ferroportin1;
diferric Tf, diferric transferrin.
Modified from Anderson and Fraser (2017) and Kondaiah et al. (2019).

Cells can take up iron in a variety of forms. All nucleated cells are capable of using transferrin-bound iron, although some cell types can take up iron in other forms, including non-transferrin-bound iron, or iron present within ferritin, heme, or Hb. Uptake and storage of iron in cells is tightly controlled by iron regulatory proteins. These ensure the iron supply to the cell is maximized when the cell has an iron deficit. Alternatively, when the cell is iron replete, the supply of iron to the cell is restricted and storage promoted. For more details of the regulation of iron homeostasis at the cellular level, see Anderson and Fraser (2017).

Whole-body iron levels are also strictly regulated. This regulation is mediated by the liver-derived peptide hepcidin which is responsible in part for the control of dietary iron absorption as well as the export of iron from body stores. When iron stores are low or iron utilization such as erythropoiesis is increased, and when plasma transferrin concentration is reduced, hepcidin production is suppressed, stimulating absorption and delivery of iron to the plasma from storage sites. As a consequence, more iron enters the circulation. In contrast, when hepatic storage iron and circulating transferrin concentrations are high, the production of hepcidin is increased, thereby reducing intestinal iron absorption and inhibiting the release of iron from body stores (Figure 17.1, Anderson and Frazer, 2017). Hepcidin production is also stimulated by cytokines such as interleukins 1 and 6 induced by inflammation (see “Acute and chronic inflammation” inSection 17.7.1 for more details).

The primary storage sites for iron are the cells of the liver, spleen, and bone marrow. Iron is stored in the form of the soluble protein complex ferritin, or insoluble hemosiderin, as noted earlier. Iron from both forms of storage iron can be mobilized efficiently when required elsewhere in the body (Anderson and Frazer, 2017). The amount of iron in the stores varies widely, depending on iron status and sex. When the iron supply exceeds the cell’s functional needs, ferritin production in the liver is increased, whereas when the supply is insufficient, iron bound to ferritin will be mobilized from the liver store.

Unlike many other micronutrients, there are no active pathways for excreting iron. Instead, mechanisms exist to maintain both cellular and whole-body iron levels within the optimal physiologic range, as noted above. Total daily iron losses are small and

occur mainly in the feces (0.6mg/d), although very small amounts are also lost in desquamated skin cells and sweat (0.2-0.3mg/d), and in urine (< 1mg/d). Most of the iron excreted in the feces is from unabsorbed dietary iron, although a small amount is systemic iron derived primarily from biliary secretions. Total daily iron losses are larger (≈ 1.3mg/d) in premenopausal women, because of the additional loss of iron in menstrual blood. On average, menstrual blood loss is 30-40mL per cycle or 0.4-0.5mg iron per day, although in some women it is much greater (Yip and Dallman, 1996).

Most of the heme iron in erythrocytes is recycled for Hb synthesis at the end of the erythrocyte's functional lifetime (on average 120d). At this time, the erythrocytes are phagocytosed by specialized macrophages in the spleen, liver, and bone marrow. The heme iron is released from catabolized Hb and is either returned to the plasma via ferroportin where it is bound to transferrin, or it is incorporated into ferritin for temporary storage, depending on the iron status of the individual. Smaller quantities of iron are exported by other cells, particularly hepatocytes. This process is termed “iron turnover”; each day 25mg systemic iron is recycled, representing 0.66% of the total iron content of the body.

17.1.3 Micronutrient interactions with the potential to affect iron absorption and/or metabolism

Deficiencies of iron, vitamin A, iodine, and zinc are major public health problems in low-income countries. They often occur concurrently, especially among infants, young children, and women of reproductive age living in impoverished settings. These high-risk groups all have high micronutrient requirements, but frequently low intakes of poorly bioavailable micronutrients. Interactions have been described between these micronutrients, although the mechanisms are not always fully understood.

Iron-vitamin A interactions. Vitamin A may affect several stages of iron metabolism, including erythropoiesis, incorporation of iron into Hb, and the mobilization of iron from ferritin stores (Zimmermann et al., 2006). Vitamin A is also said to affect iron absorption, but whether the effect is determined by vitamin A and/or iron status is uncertain (Hurrell and Egli, 2010). Reports of a positive impact of vitamin A on Hb and some iron status indices have been observed in several vitamin A supplementation / fortification trials (Michelazzo et al., 2013), although the precise mechanism remains uncertain. Based on these reports, vitamin A deficiency is now recognized as a contributor to anemia, especially in children (Calis et al., 2008).

Iron-iodine interactions. Numerous animal studies indicate that iron deficiency with or without anemia impairs thyroid metabolism (Hess et al., 2002). An iron-dependent enzyme, thyroperoxidase, acts by catalyzing the two initial steps in thyroid hormone synthesis (Hess et al., 2002). Iron deficiency may also reduce thyroid hormone synthesis by inducing alterations in the thyroid hormone feedback system (Lynch et al., 2018). In intervention studies in Morocco and the Ivory Coast, administration of both iron and iodine to iron-deficient goitrous children decreased goiter rates more effectively than did the administration of iodine alone (Zimmerman, 2002).

Iron-zinc interactions. Iron and zinc appear to compete for absorption, possibly via the shared divalent metal transporter 1 (DMT1) and/or via transporters in the apical membrane of small intestine enterocytes; the mechanism remains uncertain (Kondaiah et al., 2019). This inhibitory effect only occurs when therapeutic doses (2-3-fold molar excess or higher, iron relative to zinc) are given to humans in aqueous solutions but not in a complex food matrix (Olivares et al., 2012). To date there is no evidence that this interaction is associated with any clinically significant adverse effects (Fischer-Walker et al., 2005; Lim et al., 2013), obviating concerns regarding the use of both iron and zinc supplements in public health programs. For a comprehensive review of the possible mechanisms for this iron-zinc interaction, see Kondaiah et al. (2019).

Iron-cadmium and iron-lead. Iron deficiency is a risk factor for increased concentrations of cadmium and lead in the blood. For cadmium, this is likely due to increased absorption arising from the elevated levels of divalent metal transporter 1 (DMT) that occur in iron deficiency, but for lead, the mechanism is uncertain (EFSA, 2015).

17.1.4 Deficiency of iron in humans

Iron deficiency may arise from inadequate intakes of dietary iron, poor absorption, excessive iron losses, or a combination of these factors. Individuals at greatest risk for iron deficiency include those with high iron requirements for growth (i.e., infants, children and pregnant women), and women of child bearing age with high menstrual losses. Several risk factors have been identified. In industrialized countries, prematurity or low-birthweight, the use of a non-iron-fortified formula, the introduction in young children to cow's milk in the first year of life, and exclusive breastfeeding after 6mos without the regular feeding of iron

fortified complementary foods, are all frequently observed causal factors. Among pregnant women, diets low in iron-rich foods, a short interpregnancy interval, and increasingly, obesity, are implicated (Kemper et al., 2017). Additional contributing factors associated with an increased risk of non-anemic iron deficiency among women of child-bearing age with high menstrual losses may include low intake of flesh foods, recent blood donations, nose bleeds, and low body mass index (Heath et al., 2001).

In some impoverished settings, additional factors causing iron deficiency may include: bleeding arising from exposure to parasitic helminths (e.g., hookworm) or schistosomiasis, impaired absorption from infection / inflammation or possibly chronic *Helicobacter pylori* gastritis. In rare cases, genetic disturbances in iron homeostasis arising from genetic mutations, such as in the TMPRSS6 gene, may result in iron-refractory iron deficiency anemia (i.e., iron deficiency that is resistant to oral iron therapy) (Lynch et al., 2018).

Anemia is the most common symptom of iron deficiency. In the past, angular stomatitis, glossitis, sideropenic dysphagia, and koilonychia (spoon nails) were also considered physical manifestations of iron deficiency, although rarely observed today.

Table 17.2. Functional consequences of iron deficiency. Data from Lynch (2011).

	Consequences
Pregnancy	Increased risk of anemia and maternal morbidity Increased risk of prematurity and lower birth weight Higher infant mortality
Infants and young children	Motor and cognitive developmental delays in infancy Effects on emotional maturation Poorer academic achievement in school-age children who were iron deficient in early childhood Increased risk of severe morbidity from malaria in children < 5y Increased prevalence and duration of upper respiratory tract infections in children
All ages	Impaired physical performance and earning capacity Suboptimal response to iodine in populations with endemic goiter and increased risk of impaired thyroid function in the presence of iodine deficiency Increased risk of chronic lead poisoning in high-lead environments Increased risk of “restless leg syndrome”

These conditions probably resulted from the presence of iron deficiency along with other nutritional or environmental problems. Behavioral disturbances are still considered symptoms of iron deficiency and may include pica, characterized by the abnormal consumption of nonfood items such as dirt (geophagia) and ice (pagophagia), and restless leg syndrome, a condition of uncontrollable, uncomfortable leg movement during sleep. Table 17.2 summarizes the main functional consequences of iron deficiency during the life cycle.

Some of the less-specific physiological manifestations associated with the functional consequences of iron deficiency include fatigue, anorexia, tiredness, shortness of breath, and impaired physical performance. These are all symptoms that result from the diminished oxygen-carrying capacity of the blood. The effects of iron deficiency on impaired physical performance, notably exercise or work capacity, have been extensively reviewed by Haas and Brownlie (2001) and McClung and Murray-Kolb (2013). Based on their reviews of both animal and human studies, a profound negative effect of iron deficiency anemia on work capacity was reported; whether a comparable effect is apparent in the absence of anemia is less certain.

Numerous human studies have also investigated the association between iron deficiency and cognitive function. Interpretation of these data are difficult because neither iron deficiency nor the assessment of cognitive function outcomes have been adequately characterized in some of the studies. In a systematic review by Hermoso et al. (2011) in which only randomized controlled trials with an adequate control group were included (n=14), a modest positive effect of iron supplementation on cognition and psychomotor outcomes was reported in anemic infants and children following supplementation with iron for at least 2mos.

Nevertheless, iron is known to be essential for the normal development and function of the brain. Animal research indicates that changes in neurotransmitter homeostasis and decreases in basal ganglia are all caused by iron deficiency. Hence, early identification and treatment of iron deficiency-induced brain dysfunction appears to be critical to recover neural function and thus

prevent any long-term effects on brain development. However, such effects will depend on timing and the degree of iron deficiency. Long-term effects are most likely if the deficiency occurs during a time period when the need for iron for neurodevelopment is high (Prado and Dewey, 2014).

Numerous studies have investigated associations between maternal iron deficiency anemia or iron deficiency (without anemia) and risk of functional adverse pregnancy outcomes such as prematurity and low birthweight, and maternal and neonatal mortality. Most have been based on comparisons between the effects of daily oral supplements containing iron (usually with folic acid) versus no iron or placebo during pregnancy. Despite reducing the risk of maternal anemia and iron deficiency, the effects of iron supplementation during pregnancy on functional maternal and infant outcomes have been inconsistent. Some studies have shown significant reductions, for example, in risk of low birth weight (Imdad and Bhutta, 2012; Dibley et al., 2012), while in others, the results on maternal and infant outcomes have been less clear (Peña-Rosas et al., 2015).

Concern has been raised that some of these inconsistencies across studies may be linked to the definition of anemia applied, the timing of exposure, the pattern of changes in iron biomarkers during pregnancy, and the definitions of normal or “healthy” reference ranges for Hb applied for comparison. In an effort to address at least some of these concerns, Young et al. (2019) examined the associations of maternal Hb concentrations with a range of maternal and infant health outcomes, taking into account the timing of the Hb measurement, etiology of anemia, and Hb cutoff category. Results of their systematic review and meta-analysis confirmed that maternal Hb has an important role in relation to several maternal and child health outcomes; for more specific details for the outcomes examined, see Young et al. (2019). Nevertheless, in the studies included, very few distinguished between iron deficiency anemia and non-iron deficiency anemia in relation to maternal and child functional outcomes. Table 17.3 presents their meta-analysis summary estimates of associations of iron deficiency anemia (IDA) and non-IDA with birth outcomes: low birth-weight (LBW), small for gestational age (SGA) and preterm birth (PTB).

Table 17.3. Meta-analysis summary estimates of association of IDA and non-IDA with birth outcomes. From Young et al. (2019) The number of studies in each category is: LBW/IDA: n=2; LBW/non-IDA: n=3; SGA/IDA: n=2; SGA/non-IDA: n=4; PTB/non-IDA: n=4.

Outcome	IDA OR (95% CI)	Non-IDA OR (95% CI)
LBW	1.17 (0.95-1.43)	1.43 (0.82-2.50)
SGA	0.77 (0.68-0.87)	1.20 (0.85-1.70)
PTB		1.07 (0.68-1.70)

Note that the summary estimates for the association of low birth-weight with iron deficiency anemia and non-iron deficiency anemia were both positive, although not statistically significant, whereas for preterm birth the pooled estimate with non-iron deficient populations was non-significant, and data insufficient to derive a pooled estimate for iron deficiency anemia. The odds of small for gestational age with iron deficiency were decreased (2 studies) although there was no significant association between small for gestational age and non-iron deficiency anemia (4 studies). Given the complexity of the changes in iron homeostasis and associated iron biomarkers during pregnancy together with the diverse factors associated with etiology of anemia, research that includes some of the newer biomarkers such as serum hepcidin may help to resolve some of the reasons for the inconsistencies in the results observed.

17.1.5 Food sources and dietary intakes

High-iron foods include liver, kidney, mussels, and red meat. Foods with a medium iron content include chicken, processed meat, fish, and legumes (non-heme iron only). Milk and milk products and many fruits and vegetables are poor sources of dietary iron (USDA, 2019). Flesh foods are especially important because of their high content of bioavailable heme iron, and their enhancing effect on nonheme iron absorption (Hallberg and Rossander, 1984). Ground water may also be an important source of iron in some countries. Studies have found positive associations between consumption of ground water with a high iron content and biomarkers of iron status and Hb in Bangladesh (Merrill et al., 2011; Rahman et al., 2016).

In many industrialized countries, cereal products fortified or enriched with iron, provide the most dietary iron, followed by meat, poultry, and fish, and then vegetables and fruits (Gregory et al., 1990; McLennan and Podger, 1998; Russell et al., 1999). As an example, based on data from nine European countries, cereal products contributed more than 20% and up to 49% of the iron

intake in all population groups except infants (EFSA, 2015). In the U.S. NHANES III survey (2003-2006), about 40% of the iron in the diets of individuals > 2y was contributed by foods fortified or enriched with iron(Fulgoni et al., 2011).

Data on iron intakes are available from several nationally representative surveys. In the U.S. NHANES III survey (1988-1994), the median usual intakes of iron for adult women and men were 12mg/d, and 16-18mg/d, respectively (IOM, 2001). For adult Europeans (> 18y), average intakes ranged between 9.4 and 17.9mg/d (EFSA, 2015). In general, average iron intakes of males tend to be slightly higher than females because males generally consume larger quantities of food each day (EFSA, 2015).

The bioavailability of iron from mixed Western-style diets (for individuals with no iron stores) is estimated to range from 14-18%, provided the diets contain ample quantities of flesh foods and ascorbic acid. In low-income countries, however, dietary iron absorption is often probably only about 5-12% for individuals with no iron stores because in these plant-based diets the proportion of heme iron is low, while the content of absorption inhibitors is often significant (WHO/FAO, 2005).

The rise in the variety of food products fortified with iron has prompted research on the properties of iron fortificants. Factors affecting their choice include absorbability (i.e., the highest relative bioavailability (RBV) compared with ferrous sulfate), absence of unacceptable changes to the sensory properties (i.e., taste, color, texture) of the food vehicle, and the cost of the fortificant. The iron compounds used for fortification or for enrichment can be divided into three categories — water soluble; poorly water soluble but soluble in dilute acid; water insoluble and poorly soluble in dilute acid. Almost all enter the common nonheme iron pool in the gastrointestinal tract and are absorbed like native nonheme iron compounds. The third category, however, — water insoluble and poorly soluble in dilute acid — do not fully enter the common nonheme iron pool, and their absorption is poor.

The compounds currently used as iron fortificants in order of frequency of use are : (1) ferrous sulfate; (2) ferrous fumarate; (3) encapsulated ferrous sulfate or encapsulated ferrous fumarate; (4) electrolytic iron (a pure form of iron powder with a small particle size) or ferric pyrophosphate; (5) sodium-iron-ethylene-diamine-tetra-acetic-acid (NaFeEDTA); and (6) iron amino acid chelates, particularly iron-glycinate chelates (Lynch et al., 2018). The newer iron fortificants such as NaFeEDTA and ferrous bisglycinate provide better protection against nonheme iron absorption inhibitors. Sodium iron EDTA is used in China to fortify wheat products, whereas ferrous bisglycinate, although expensive, is especially suited to the fortification of liquid whole milk and dairy products. For a detailed review of iron fortificants, see WHO (2006).

17.1.6 Effects of high intakes of iron

Cases of acute iron toxicity have been reported, mainly among children who accidentally ingest medicinal iron supplements (IOM, 2001). More common are adverse gastrointestinal effects including nausea, vomiting and diarrhea following the administration of high doses of iron supplements, especially when they are taken without food (Brock et al., 1985). Excessively high intakes of iron supplements may induce copper deficiency by an adverse effect on copper absorption, although this effect may only occur when copper status is low or marginal (Cohen et al., 1985). High intakes of iron supplements taken without food may reduce zinc absorption (Sandström et al., 1985), but generally not when iron supplements or iron fortificants, such as “sprinkles”, are taken with food (Olivares et al., 2012; Esamai et al., 2014; Davidsson et al., 1995), as noted earlier. In anemic Cambodian women, however, a daily multi-micronutrient supplement containing 60mg of iron consumed with food blunted the increase in serum zinc observed among the women receiving the same supplement without iron; no data on zinc absorption were collected (Holmes et al., 2019).

Possible toxic effects of rising iron stores over time have been implicated in the pathogenesis of several common diseases of aging. However, evidence for a causal relationship in the general population between the level of dietary iron intake or body iron content and risk for cancer (Turner and Lloyd, 2017) and coronary heart disease (Hunnicutt and Xun, 2014) remains inconclusive. Moreover, there is no evidence that dietary iron is associated with arthritis, diabetes mellitus, or neurodegenerative disease (EFSA, 2015).

Certain genetic mutations can cause iron overload. One of the most well-known examples is hereditary hemochromatosis arising from a mutation in HFE gene that results in a reduced or negligible expression of hepcidin. This results in an inability to limit absorption of dietary iron, so iron accumulates in many organs. Hemochromatosis is relatively common in Northern European populations (Beutler et al., 2002); heterozygotes for the mutation associated with hemochromatosis, however, do not appear to be at an increased risk of iron overload.

Iron overload has also been associated with other genetic defects such as atransferrinemia, aceruloplasminemia, and Friedreich

ataxia. African iron overload, previously termed “Bantu cirrhosis”, is also thought to have a genetic component in addition to the excessive intakes of iron from food and beer prepared in iron utensils (EFSA, 2015). Iron overload may also result from the multiple blood transfusions used to treat certain inherited disorders such as iron-loading anemia (e.g., β-thalassemia). In most iron overload syndromes, iron is sequestered in ferritin and hemosiderin in all tissues throughout the body.

Concern has been raised that supplementation of iron given in non-physiological amounts can increase the risk of bacterial and protozoal infections (see “malaria” in Section 17.11.1 for more details). As approximately 40% of the population in the world is exposed to malaria and it is endemic in over 100 countries, resolving this concern is essential. Neuberger et al. (2016) reviewed the available evidence based on a meta-analysis of 35 randomized controlled trials (31,955 children). They concluded that oral iron supplementation does not increase the risk of clinical malaria provided regular malaria prevention or management services are provided. Therefore, in resource-poor settings, routine iron supplementation can be given to children without screening for anemia or for iron deficiency provided regular malaria surveillance and treatment services are provided. However, such services for malaria are not always available in low-income countries. For a review of the safety and benefits of iron supplements for pregnant women in low-income countries, consult Mwangi et al. (2017).

Recommendations by the World Health Organization (WHO) on the prevention and control of iron deficiency in infants and young children in malaria-endemic areas are available (WHO, 2007). Instead of using iron supplements, a safer strategy for improving iron status may be the use of lower quantities of iron provided within a food matrix (i.e., fortified food); see Prentice et al. (2017) for more details.

The U.S. Food and Nutrition Board has set the Tolerable Upper Intake Level (UL) based on the gastrointestinal side effects associated with high levels of iron consumed on an empty stomach. The UL for iron for adults, including pregnant and lactating women ≥ 19y is 45mg/d, whereas the level for infants and children 1-18y is 40mg/d (IOM, 2001). No UL has been set for iron by the European Food Safety Authority — EFSA (2006; 2015) or by WHO/FAO (2005).

17.1.7 Biomarkers of exposure

To estimate exposure to the intake of dietary iron, conventional dietary assessment methods must be used. The method selected depends on several factors, including the study objective, the setting, and available resources; for more details, see Chapter 3. Estimates of habitual usual iron intakes are needed in view of the length of time required by iron status biomarkers to reach a steady state in response to changes in dietary intakes. Usual iron intakes are also essential to estimate the prevalence of inadequate intakes for population groups and for the investigation of associations between dietary intakes and health outcomes. Details of the statistical adjustments for converting observed iron intakes to usual iron intakes can also be found in Chapter 3. Alternatively, for large-scale population studies, a food frequency questionnaire (FFQ) that is designed to assess intakes of iron-rich foods and iron-absorption modifiers over a longer time period may be used, provided the FFQ has been validated for the study setting (see Chapter 3, Section 3.1.6).

As noted earlier, inadequate intakes of poorly available dietary iron are a major concern in the plant-based diets of many low-income countries. Bioavailability is defined as the extent to which the iron is absorbed from the diet and used for normal bodily functions (Lynch et al., 2018). Several algorithms have been developed to predict the bioavailability of iron based on dietary data that include the form of the iron (i.e., heme or nonheme) and the content of known absorption modifiers in a meal. Of the algorithms available, the model of Hallberg and Hulthén (2000) is the most detailed, taking into account the effects of all known modifiers of nonheme iron absorption as well as interactions, with adjustments for the iron status of the individual. However, as with all the other algorithms developed, the model of Hallberg and Hulthén only predicts nonheme iron absorption so assumptions about the intake and absorption of heme iron must be made. Furthermore, the model is difficult to apply, in part because of incomplete food composition data for the absorption modifiers. Indeed, all algorithms have limitations and often underestimate bioavailability, with a 3-fold variation in estimates when using different algorithms (Beard et al., 2007).

In view of the limitations of the algorithms to estimate iron bioavailability, FAO/WHO (2002) have developed a qualitative approach to estimate iron absorption based on data on iron absorption from typical meals in Asia, India, Latin America, and Western countries. Applying the FAO/WHO approach, diets can be categorized as having a low, intermediate, or high bioavailability based on the major food sources of iron, as shown in Box 17.1. The estimates of absorption given in Box 17.1 refer to non-anemic persons with normal iron transport (i.e., with normal Hb) with no iron stores. When individuals have iron deficiency anemia, absorption may be increased by 50% (i.e., increasing to percentages of 7.5, 15, and 22.5, respectively, for the low-, intermediate-, and high-bioavailability diets). Note that for Western-type diets, FAO and WHO now propose two categories of bioavailability —

12% and 15% — depending on the meat content of the diet; see Gibson and Ferguson (2008) for further details.

Box 17.1. Using the FAO/WHO qualitative approach to estimate iron bioavailability

Uses the data on major food sources to estimate iron bioavailability

- **Low-bioavailability diet** (iron absorption of about 5%): a simple, monotonous diet containing cereals, roots, and/or tubers, and negligible quantities of meat, fish, or ascorbic acid-rich foods;
- **Intermediate-bioavailability diet** (iron absorption of about 10 percent): consists mainly of cereals, roots, and/or tubers, and minimal quantities of food of animal origin and ascorbic acid, both of which promote iron availability;
- **High-bioavailability diet** (iron absorption of about 12 or 15%): a diversified diet containing generous quantities of meat, poultry, fish and/or foods containing high amounts of ascorbic acid.

Modified from FAO/WHO (2002).

Recently a new approach to calculate dietary iron bioavailability has been developed using a probability model. The approach is based on calculated iron requirements, data on daily total iron intake, and the distribution of serum ferritin values in the population sample; total iron absorption at any level of iron status can be estimated (Dainty et al., 2014). Unlike the earlier algorithms, this new approach does not require data on iron absorption modifiers or the heme content of the diets.

Using the approach described above, data from a representative group of adults from the U.K. National Diet and Nutrition Survey were used to estimate total iron absorption from a mixed Western style diet, using a target serum ferritin concentration of 30µg/L for iron storage status. At this level, estimates for the predicted iron absorption for men and premenopausal women were 16% and 18%, respectively (EFSA, 2015). Note an interactive tool is available (Fairweather-Tait et al., 2017) that could be adapted for use in populations consuming habitual plant-based diets with low-iron bioavailability, provided the appropriate correction for the impact of inflammation on serum ferritin concentrations is performed, when necessary (see “Acute and chronic inflammation” in Section 17.7.1 for more details). Nevertheless, because both the iron requirements and dietary iron must be in a steady state for at least a year when applying this approach, it is not suitable for certain groups such as children, pregnant women, or immediately after the onset of menopause.

17.1.8 Interpretation of iron intakes

The final stage of the assessment of exposure to the intake of iron is to evaluate the usual intakes of the population or an individual in relation to the nutrient reference values (NRVs); see Chapter 8a for details of the derivation of the NRVs. To estimate the physiological requirements for iron, the factorial approach, based on estimates of the quantity of absorbed iron needed to replace iron losses, is used by most expert groups, including WHO/FAO (2005), IOM (2001), and EFSA (2015). Next, adjustments are made to the estimates of physiological requirements to yield a dietary iron requirement (i.e., Average Requirement; AR) by taking into account the bioavailability of iron in the habitual diet. The chosen bioavailability factors (see Chapter 8a, Section 8a.8 for more details) are partly responsible for differences in the dietary iron requirements (i.e., ARs) across countries. Therefore, care must be taken to ensure that the dietary iron requirements chosen for the evaluation of iron intakes are appropriate for the habitual diet of the population under study.

Some expert groups (e.g., IOM, 2001; EFSA, 2015) apply a fixed bioavailability factor for iron to estimate the dietary requirement (AR), even though the efficiency of absorption may vary with life-stage and health status. Others, including WHO/FAO (2005), do not present average dietary requirements (ARs) for iron, instead providing only estimates of Recommended Intakes (RIs) from which the ARs for dietary iron can be derived. These calculated AR values are available in WHO (2006) and are the values that should be used to evaluate the prevalence of inadequate intakes of population groups in a low-income setting. In the higher-income countries, however, the AR values set by IOM (2001) or EFSA (2015) may be appropriate. The full-probability approach must be applied to estimate the prevalence of inadequate intake for iron for certain population subgroups, notably children(1-8y), menstruating adolescents (14-18y) and adult women, because their iron requirements are not symmetrical about the AR. For the details of the application of the full probability approach both manually and using appropriate software, see Chapter 8b. For the manual calculation, the reader is also advised to consult Gibson and Ferguson (2008).

To estimate iron bioavailability in the diet of an individual, the qualitative model of FAO/WHO (2002) (Box 17.1.) that takes into account the major food sources of iron and the presence of potential absorption modifiers can be used. Alternatively, bioavail-

ability can be calculated from an appropriate iron algorithm. Inferences can then be made about the adequacy of the iron intake of the individual by comparing the difference between the estimate of the usual iron intake and the corresponding selected values for the AR and RI, taking the assumed bioavailability into account. For details on the guidelines for the qualitative interpretation of individual iron intakes, see Chapter 8b.

17.1.9 Introduction: biochemical biomarkers of iron status

In the past, iron deficiency has been assumed to cause 50% of all cases of anemia worldwide. The proportion is now thought to be much lower. In a systematic review, Petry et al. (2016) estimated that only 25% of the anemia in pre-school children, and 39% in non-pregnant women of reproductive age, is attributable to iron deficiency. In their review, unlike in many earlier studies, the probable exposure of a country to infection and chronic inflammatory conditions was taken into account. Their findings highlighted the challenge of distinguishing between anemia due to iron deficiency and that due to inflammation (i.e., anemia of chronic disease). This is a concern in low-income countries and in the elderly who are also at high risk for inflammation (see “Acute and chronic inflammation” in Section 17.7.1 for more details).

The gold standard test for the diagnosis of iron deficiency anemia is a stained bone marrow aspirate or biopsy. This provides a semi-quantitative estimate of the size of the body iron store. However, this test is invasive and costly, and requires an experienced person to evaluate the stained specimen, and so is not commonly used (Lynch et al., 2018). Instead, several hematological parameters are used to diagnose anemia, along with serum-based iron biomarkers and biomarkers of inflammation to identify iron deficiency or iron deficiency anemia; these are discussed in detail in the following sections. For a summary table of each of these measures, and their advantages and limitations, the reader is referred to Lynch et al. (2018).Three stages in the development of iron deficiency anemia can be recognized and are best characterized by the use of multiple biomarkers (Section 17.10). The three stages are the following:

Iron depletion, the first stage, is characterized by a progressive reduction in the amount of storage iron in the liver. At this stage, the supply of iron to the functional compartment is not compromised so levels of transport iron and hemoglobin are normal. However, the progressive depletion of iron stores will be reflected by a fall in serum ferritin concentrations.

Iron-deficient erythropoiesis, the second stage, is characterized by the exhaustion of iron stores and is also referred to as “iron deficiency without anemia”. At this stage the iron supply to the erythropoietic cells is progressively reduced and decreases in transferrin saturation occur (Section 17.6). At the same time, there are increases in serum transferrin receptor (Section 17.9) and erythrocyte protoporphyrin concentrations (Section 17.8). Hemoglobin levels may decline slightly at this stage, although they usually remain within the normal range.

Iron-deficiency anemia, the third and final stage of iron deficiency, is characterized by the exhaustion of iron stores, declining levels of circulating iron, and presence of frank microcytic, hypochromic anemia (i.e., red blood cells are small and contain less Hb than normal). The main feature of this stage is a reduction in Hb (Section 17.2), arising from the restriction of iron supply to the bone marrow. Decreases in the hematocrit (Section 17.3) and red-cell indices also occur (Section 17.4). Examination of a stained blood film allows confirmation of the presence of hypochromia (a lower Hb than normal) and microcytosis (abnormally small red-blood cells).

This page titled 14.1: Introduction and functions of iron (17.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.2: Hemoglobin (17.2)

Iron is an essential component of the Hb molecule, the oxygen-carrying pigment of red blood cells. Each Hb molecule is a conjugate of a protein (globin) and four molecules of heme, as noted earlier. Anemia develops when circulating red blood cells are insufficient to meet physiological oxygen-carrying needs, and is identified by a Hb concentration below a defined cutoff-value. Measurement of the Hb concentration in whole blood is probably the most widely used test for iron-deficiency anemia, even though only a subset of anemic individuals in a population will have iron deficiency anemia. Anemia may also be due to malaria, infection, hemoglobinopathies, or deficiency of several other micronutrients besides iron.

Table 17.4. Within-subject, day-to-day coefficient of variation (CV%) for Hb, serum ferritin, and serum iron for "healthy" subjects. M, r

Reference	Hemoglobin	Serumferritin
Dawkins et al. (1979)	—	15 (M, F)
Gallagher et al. (1989)	1.6 (F)	15 (F)
Statland and Winkel (1977)	—	—
Statland et al. (1976)	—	—
Statland et al. (1978)	3 (M, F)	—
Pilon et al. (1981)	—	15 (M, F)
Romslo and Talstad (1988)	—	13 (M, F)
Borel et al. (1991)	4 (M, F)	14 (M)
Borel et al. (1991)	4 (F)	26 (F)

Hence, the specificity of Hb for iron deficiency anemia is poor. Hemoglobin is also an insensitive measure of iron deficiency because it only falls during the third stage of iron deficiency, when iron stores are exhausted and the supply of iron to the tissues is compromised. In addition, considerable overlap exists in the Hb values of normal nonanemic and iron-deficient individuals, further reducing its sensitivity for screening for iron deficiency. Consequently, Hb should not be used as the only measure of iron deficiency anemia in individuals, except in very specific circumstances (Section 17.2.4). Limitations of using Hb as a measure of iron deficiency are discussed below.

17.2.1 Factors affecting Hb concentrations

Daily biological variation in Hb concentrations is significant (Van Wyck et al., 2011; Sennels et al., 2011). Hemoglobin values tend to be lower in the evening than in the morning, by amounts of up to 10g/L. Within-subject day-to-day variation is low (CV≈3%) (Table 17.4).

Age is an important determinant, particularly for infants and children. Hemoglobin concentrations are higher (130-180g/L) at birth (Lorenz et al., 2013) than any other time in life, reflecting fetal adaption to the oxygen- deficient environment of the uterus. After birth, Hb concentrations fall, reaching their lowest point at about 2mos This fall is probably due to the sudden decrease in erythropoiesis arising from the increase in oxygen delivery to the tissues (Zierk et al., 2015). (Figure 17.2).

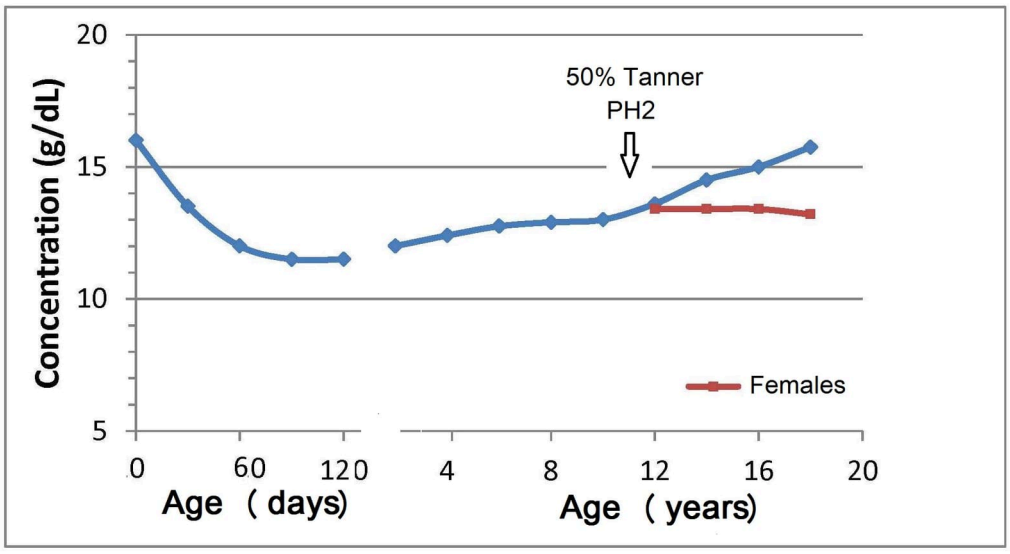


Figure 17.2. Changes in median concentration Hb with age. Arrow shows the time point when 50% of the subjects reach Tanner stage PH2 and sex differences become significant. Redrawn from Zierk et al.(2015).

Thereafter, Hb concentrations gradually rise (Domellöf et al., 2002a), with a marked increase at puberty in males as a result of accelerated growth (Figure 17.2). The arrow shows the point when 50% of the population reach Tanner stage PH2 (i.e., the appearance of pubic hair), emphasizing the sex-dependent changes after puberty (Zierk et al., 2015). Around the fifth decade of life, Hb levels decline. Several factors have been implicated in this decline, including the development of anemia of chronic disease with aging.

Sex differences in Hb are apparent at 6mos, with boys having slightly lower Hb concentrations than girls. Such discrepancies appear to be greater in infants with a birth weight < 3500g (Domellöf et al., 2002). By the second decade of life, however, females have values that change very little after 12y (Table 17.5). In young adults, the Hb concentration for men is on average about 20g/L higher than for women, due to testosterone which results in both a larger body size and larger erythrocyte mass. For women, the trend for lower Hb values is due to chronic menstrual blood loss. These sex-related differences, although diminishing gradually with increasing age, remain (Table 17.5).

Table 17.5. Mean (SD) hemoglobin concentrations (g/L) by age and sex. Data from National Health and Nutrition Examination Survey (1988-1994).Individuals of all races with abnormal or missing values for transferrin saturation, erythrocyte protoporphyrin, serum fer mean cell volume were excluded. From Looker et al. (1997).

Age (y)	Females	Males
1-2	122.0 (7.34)	
3-5	124.4 (7.57)	
6-11	130.9 (7.92)	
12-15	134.3 (9.27)	142.4 (10.0)
16-19	133.7 (8.21)	152.9 (10.03)
20-49	134.8 (9.12)	153.0 (9.68)
50-69	136.5 (9.82)	150.1 (10.64)

Age (y)	Females	Males
≥ 70	135.6 (10.68)	145.3 (12.87)

Ethnicity is known to influence Hb concentrations. Individuals of African descent in the United States have Hb values 5-10g/L lower than Caucasians, irrespective of age, income, or iron deficiency. It is likely that a genetic factor is involved. In contrast, Hb concentrations for other U.S-based races are similar to those of U.S Caucasians (Nestel, 2002; Cheng et al., 2004). In 2001, WHO / UNICEF / UNU (2001)recommended adjusting Hb concentrations downward by 10g/L for individuals of African descent, irrespective of age. This recommendation has now been abandoned in view of the paucity of data related to ethnicity. Instead, standard cutoffs for Hb, irrespective of race, are now recommended (WHO, 2011).

Cigarette smoking is associated with higher concentrations of Hb (0.3-0.7g/L) in adults. This is attributed to a reduction in the oxygen-carrying capacity of the blood arising from the carbon monoxide-induced increase in carboxyhemoglobin levels, which has no oxygen-carrying capacity. An adjustment to define anemia for smokers has been defined by the U.S. CDC based on U.S. NHANES II data (see Table 17.6). The adjustments should be subtracted from the measured values. These adjustments are additive so that if smokers are living at higher altitudes (see below), two adjustments would be needed.

Higher altitudes generate an adaptive response to the lower partial pressure of oxygen and the reduced oxygen saturation of blood.

Table 17.6. Altitude and smoking adjustments to measured hemoglobin concentrations. From Nestel(2002) and WHO (2011): Hemog concentrations for the diagnosis of anemia and assessment of severity. Vitamin and Mineral Nutrition Information System, Geneva.

Altitude (metres above sea level)	Measured Hb adjustment (g/L)
< 1000	0
1000	– 2
1500	– 5
2000	– 8
2500	– 13
3000	– 19
3500	– 27
4000	– 35
4500	– 45
Smoking status	adjustment (g/L)
Non-smoker	0
Smoker (all)	–3
0.5-1 packet/d	–3
1-2 packets/d	–5
≥ 2 packets/d	–7

As a result, red blood cell production increases to ensure sufficient oxygen is supplied to the tissues (Hurtado et al., 1945). This adaptive response becomes significant at elevations above 1000m, when adjustments for altitude for Hb cutoffs published by the CDC should be applied (CDC, 1989). The changes with altitude are curvilinear, with the increase in Hb concentration becoming more marked as altitude increases (Table 17.6). Even individuals living at altitudes with Hb concentrations that are above the cutoff said to be indicative of anemia, but with concomitant iron deficiency, may respond to iron supplementation by an increase in Hb (Berger et al., 1997).

Iron-deficiency anemia develops during the third stage of iron deficiency. At this stage, iron stores are exhausted, the supply of iron to support the production of red blood cells is compromised, and as a consequence Hb concentration decreases. Iron-deficiency anemia is characterized by microcytic, hypochromic anemia. Considerable overlap exists in the Hb values of normal nonanemic and iron-deficient individuals, noted earlier, which further decreases the sensitivity of Hb measurements (WHO, 2017).

Pregnancy results in an expansion of both plasma volume and the red-cell volume. Near the end of the second trimester of pregnancy, the increase in plasma volume is not matched by a proportionate increase in red cells, so Hb becomes even more diluted. In women who are not taking iron supplements, the decrease in Hb reaches a low point at about 28-36wks (on average, ≈ 2g/dL lower than pre-pregnancy Hb levels) (Figure 17.3). Among iron-supplemented women, Hb concentrations are reportedly about 1g/dL higher at term compared with those unsupplemented during pregnancy (Fisher and Nemeth, 2017).

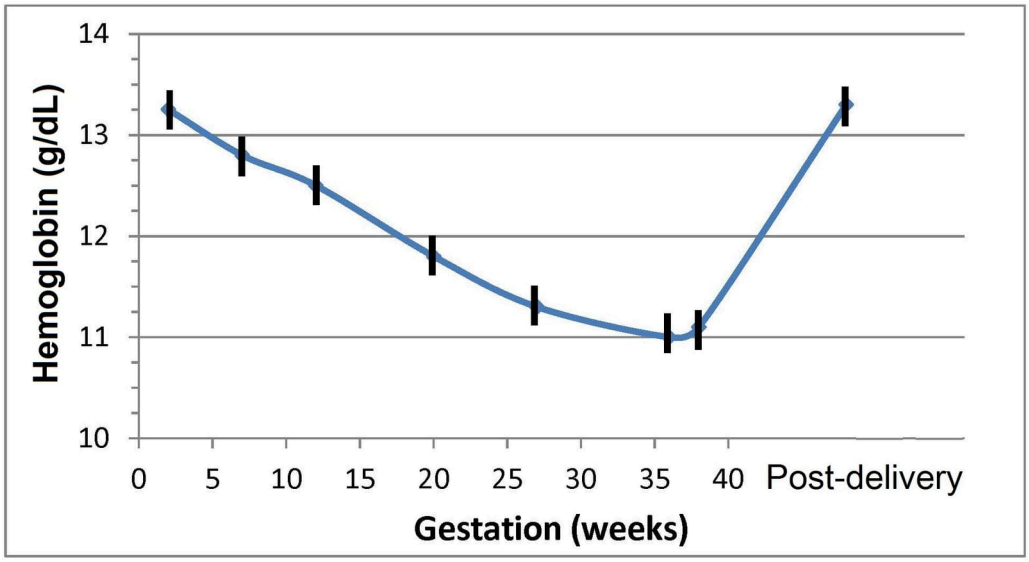


Figure 17.3. Mean hemoglobin concentrations (n=60) during normal unsupplemented pregnancy. Redrawn from Fisher and Nemeth (2017).

Other micronutrient deficiencies besides iron are associated with anemia and, by definition, with low Hb concentrations. These include deficiencies of vitamin A, B_6, and B_{12}, riboflavin, folic acid, and copper; details of the role of vitamins in the etiology of anemia are given in Fishman et al. (2000), and for copper in Myint et al. (2018). More recently, zinc, selenium, and vitamin D deficiency have also been implicated in anemia(Gibson et al., 2008; Smith & Tangpricha 2015; Houghton et al., 2016). Details of the association and the proposed mechanisms linking zinc with Hb concentrations in both preschool children and women of reproductive age are reported in Greffeuille et al.(2021). For selenium, reductions in the activity of two selenoenzymes, glutathione peroxidase and thioredoxin reductase, have been linked with anemia (Houghton et al., 2016). There is emerging evidence that vitamin D may be protective against anemia by supporting erythropoiesis. For more details of proposed mechanisms, see Smith and Tangpricha (2015).

Blood donations, when regular and repeated, may cause a decrease in body iron stores (Heath et al., 2001), leading to a fall in Hb concentrations and the development of iron deficiency anemia in some individuals. This is not surprising as 1mL of blood contains 0.5mg of iron, so a single donation of 450mL blood will contain nearly 250mg of iron. Therefore, if blood is donated thrice yearly, nearly 1g iron annually will be lost. Currently, iron supplementation for blood donors is often not a standard care. Hence, the finding that for regular blood donors in India, for example, body iron stores were inversely proportional to the lifetime number of donations, is not unexpected (Reddy et al., 2020).

Anemia of chronic disease (ACD) is a normochromic, normocytic anemia (i.e., the average size and Hb content of red cells are within normal limits) with characteristically mild (90g/L) to moderate (80g/L) Hb concentrations. It can result from a variety of different conditions. Acute and chronic infections such as malaria, tuberculosis, bacterial infection and HIV/AIDS as well as inflammatory conditions such as rheumatoid arthritis or malignancy can all lead to the anemia of chronic disease. All these conditions induce the acute-phase response whereby pro-inflammatory cytokines (e.g., IL-6 and leptin) stimulate the release of

hepatic hepcidin into the circulation. Hepcidin in turn, blocks the release of iron into the circulation from storage sites (e.g., reticulo-endothelial cells) independent of iron status, resulting in an increase in serum ferritin. At the same time, hepcidin inhibits iron absorption. As a consequence, serum iron levels fall, the iron available for erythropoiesis is reduced, and Hb concentrations decline (Anderson and Frazer, 2017). In chronic inflammation, erythropoietin production may also be impaired and red cell survival diminished (Raiten et al., 2015).

Malaria parasites cause low Hb concentrations and thus anemia by: (a) increased hemolysis with erythrocyte destruction of both infected and non-infected red cells; (b) suppression of red blood cell formation (i.e., suppressed erythropoiesis) by the action of hemozoin, a product of Hb catabolized by the parasites; and (c) induction of the acute phase response and the subsequent upregulation of hepcidin. The latter inhibits both the release of iron from storage sites into the circulation and the absorption of dietary iron, thus also blocking erythropoiesis (Ghosh and Ghosh, 2007; Drakesmith and Prentice, 2012).

Malaria is the most significant parasitic disease in humans and one of the primary causes of anemia globally. It is caused through infection by parasites of the *Plasmodium* genus, with the *P. falciparum* causing the most serious consequences. Groups at highest risk are pregnant women and children, especially younger children (i.e., < 72mos) with high levels of storage iron (Barffour et al., 2017).

Intestinal parasitic infections such as Helminths (e.g., hookworms: *Necator americanus* and *Ancylostoma duodenale*) may cause low Hb values and subsequent anemia due to chronic gastrointestinal blood loss. Hookworms become attached to the mucosa of the upper small intestine and feed on blood which is subsequently expelled through the intestinal tract of the hook worm. Secondary blood loss may also occur from bleeding of the damaged intestinal mucosa (Stoltzfus and Dreyfuss, 1998). The magnitude of the blood loss is proportional to the burden of hookworm infection in an individual (WHO, 2017). Infection with hookworm is treated easily with antihelminthic medication such as Albendazole.

Schistosomiasis is caused by the blood-dwelling worm of the genus *Schistosoma*, and leads to blood loss, and low Hb concentrations. The precise mechanism is unclear, although splenic sequestration of erythrocytes, increased hemolysis, or the anemia of chronic disease may play a role (WHO, 2017). Infection occurs through exposure to contaminated fresh water, primarily in Sub-Saharan Africa where this devastating parasitic disease is widespread. All mature species of *Schistosoma* can be treated with the drug Praziquantel (Gryseels, 2012).

***Helicobacter pylori* infection** is a gram-negative bacterium that can be acquired by means of oral-oral or fecal-oral transmission, the latter often via ingestion of contaminated water. The pathogen colonizes the gastric epithelium, surviving the highly acidic lumen. Most individuals with *H. pylori*remain asymptomatic. However, infection predisposes the development of diseases such as peptic ulcers, gastric cancers, and lymphoma, all of which can cause chronic bleeding, resulting in iron deficiency anemia (de Brito et al., 2019).

Genetic hemoglobin disorders can cause low Hb concentrations and anemia. They are classified as thalassemias when there is impaired synthesis of Hb chains. Alternatively, in structural Hb variants (i.e. hemoglobinopathies), defective Hb chains result from genetic alterations in the physical structure of Hb. The clinical features of genetic Hb disorders vary with the severity of the genetic defect and their effect on the structure and function of Hb. Some present as mild-to-severe anemia, whereas others are associated with multiple clinical complications that may sometimes be severe enough to cause death *in utero* (Weatherall, 2010; Weatherall and Clegg, 2001). Certain serum biomarkers (e.g., ferritin and transferrin receptor) are affected, limiting their usefulness for detecting iron deficiency where these disorders are widespread; see “Genetic Hb disorders” in both Section 17.7.1 and Section 17.9.1 for more details.

In a recent study of the global burden of anemia, cases of anemia in 11.6% of females and 9.9% of males were said to be caused by genetic Hb disorders (Kassebaum et al., 2016), the most common being α- and β-thalassemia, together with sickle cell anemia, Hb C and Hb E. Heterozygous carriers of these diseases may have mild or no abnormalities (Barrera-Reyes and Tejero, 2019). These genetic Hb disorders are highly prevalent in areas of the world where there is a high burden of malaria as even the heterozygous conditions appear to confer some protection against severe malaria.

Genetic disorders of red-cell enzymes are also associated with the risk of anemia, the most frequent being glucose-6-phosphate dehydrogenase deficiency (G6PD). This disorder renders red-blood cells susceptible to chronic, or acute intermittent, hemolytic anemia under some conditions that induce oxidative stress (e.g., certain medications, during infections, exposure to certain foods). Like some genetic Hb disorders, G6PD is also common in malaria-endemic regions, with the highest prevalence reported in Sub-Saharan Africa (Barrera-Reyes and Tejero, 2019).

Certain disease states such as HIV and tuberculosis can reduce Hb concentrations. In HIV infection, low Hb values may

arise from indirect effects arising from opportunistic infections, nutritional deficiencies, and the negative effect of antiretroviral therapy. Direct effects of the HIV virus include an effect on hematopoietic progenitor cells and a reduced responsiveness to erythropoietin (WHO, 2017).

In tuberculosis, low Hb values arise from increased blood loss, hemolysis (blood in sputum), and decreased production of erythrocytes. Poor appetite and thus reduced food intake, are additional contributing factors which in turn, may result in inadequate intakes of multiple micronutrients (WHO, 2017).

Both HIV infection and tuberculosis are also linked to the anemia of chronic disease, as noted above. For discussion of the impact of anemia of chronic disease on ferritin concentrations; see "Acute and chronic liver disease" in Section 17.7.1.

17.2.2 Interpretive criteria — use of the distribution of hemoglobin reference values

At the population level, Hb concentrations can be evaluated by comparison with a distribution of Hb reference values using percentiles. The reference values are obtained from a cross-sectional analysis of a reference sample group, which in most cases consists of an apparently "healthy" population sampled during a nationally representative survey and assumed to be disease-free. Consequently, some of the participants in the sample will have Hb concentrations lower than normal because they are affected by some of the factors, discussed above, known to influence Hb concentrations. Nevertheless, such reference distributions are often used as a standard for comparison with the Hb distribution measured in a study population. Examples of the reference distributions of Hb concentrations compiled in this way are shown below:

- U.S. Tables of mean, SEM, selected percentiles (5th, 10th, 25th, 50th, 75th, 90th, 95th) for Hb (g/L) by age and sex for persons randomly selected to participate in the U.S. NHANES III (1988-1994) population. See Appendix 17.1 (IOM, 2001). Children fed human milk and pregnant and lactating women were excluded, but not potentially iron deficient individuals.
- Results from two of the earlier U.K. Diet and Nutrition Surveys (Gregory et al., 1990; Gregory et al., 2000; present the mean, median, and lower and upper 2.5 or 5th percentiles for Hb for adults and older children respectively. Information dealing with UK National Survey years 9-11 (2016-2017 and 2018-2019) includes methodological details. A condensed ("zipped") comprehensive set of data tables based on information from the U.K. National Survey (years 1-9) includes more recent data on Hb.

In some cases, the distribution of Hb concentrations has been drawn from a "healthy" reference sample in which participants with conditions known to affect Hb concentrations have been excluded. Examples include:

- U.S. Tables of mean, SEM, selected percentiles (5th, 10th, 25th, 50th, 75th, 90th, 95th) for Hb (g/L) by age and sex (but not race) for a "healthy" reference sample based on data from the U.S. NHANES III 1988-1991 survey in which those with a higher risk of iron deficiency and pregnant women were excluded. Exclusion criteria were based on an abnormal value or missing value for 3 of the 4 tests (free erythrocyte protoporphyrin, transferrin saturation, serum ferritin, and mean cell volume (MCV)) (Looker et al., 1997).
- U.S. Tables of mean, SEM, selected percentiles (2.5^{th}, 5th, 50th, 80th, 95th, 97.5^{th} for Hb (g/L) for a "healthy" reference sample from U.S. NHANES III by sex for nine age groups ranging from 10y-75+y for three major racial categories (Mexican Americans, non-Hispanic black, and non-Hispanic white). Several exclusion criteria were applied; see Cheng et al. (2004) for further details and additional plots. The same team of investigators have also compiled reference distributions for Hb (g/L) by waist circumference according to sex and the same three major U.S. racial categories for individuals 25-55y based on U.S. NHANES III data from 2005-2010 (Vuong et al., 2014).
- International Tables of mean, median, gestational age-specific smoothed percentiles (3rd, 10th, 50th, 90th, 95th, 97th) for maternal Hb (g/L) for a "healthy" reference sample of pregnant women who received optimal antenatal care (Ohuma et al., 2020). The actual values for the percentiles for maternal Hb from 14-40wk gestational age are presented in Appendix 17.2. They are derived from prospective data on healthy, well-nourished women (n=3502) living in eight countries whose healthy babies were followed up to 2y. Equations are also provided to allow calculation of any desired percentiles for maternal Hb according to gestational age in weeks; see Ohuma et al. (2020) for more details. Note when gestational age is unknown, then

the median value for Hb (g/L) of the maternal Hb percentiles should be used. These median values are: 98, 101, 104, 118, 131, 133, and 137g/L for the 3rd, 5th, 10th, 50th, 90th, 95th, and 97th percentiles, respectively.

17.2.3 Hemoglobin cutoff values

Individuals in a population can be classified as "at risk" by comparing Hb concentrations with either statistically predetermined reference limits drawn from the reference distribution, or clinically or functionally defined "cutoff values". For Hb, where a single reference limit has been defined, in most cases this has been set at the 5th percentile of a "healthy" reference distribution or one in which the population sampled are assumed to be disease free, as noted in Section 17.2.2. Technically, a statistically defined reference limit is not a "cutoff value" because it has not been linked with physiological or health outcomes, so these two terms should not be used interchangeably. Nevertheless, invariably, the term "cutoff values" is applied irrespective of the method used to define them. For more details, see Chapter 15: Evaluation of nutritional biomarkers.

Table 17.7 compares the Hb "cutoff" values (g/L) set in 1968 by WHO (WHO, 1968) and by Looker et al. (1997) for the United States. The WHO cutoffs were based initially on data from small studies of predominantly Caucasian Europeans and Canadians, whereas the U.S. Hb data were from the NHANES III survey (1988-1991). In both cases, the cutoffs set are based on the 5th percentile values of the reference distribution, although the latter compiled by Looker et al. (1997) comprised a "healthy" reference sample because those with conditions known to affect Hb concentrations were excluded, as noted in Section 17.2.2. Technically because these "cutoff-values" were statistically derived rather than being linked with physiological or health outcomes, they both should be referred to as "reference limits", as noted above.

Concern has been raised over the appropriateness of these statistically derived reference limits for defining anemia among certain age groups and ethnicities (Johnson-Spear and Yip, 1994; Beutler and Waalen, 2006; Jorgensen et al., 2019). As a result, Addo et al. (2021) have used cross-sectional data from a multinational sample across 27 nutrition surveys from 25 countries to re-examine the appropriateness of using the 5th percentile WHO (1968) Hb cutoffs across countries for preschool children and nonpregnant women. Only data from "healthy" individuals were included, so persons with iron deficiency (ferritin < 12µg/L for children or < 15µg/L for women), vitamin A deficiency (serum retinol-binding protein or retinol < 0.70µmol/L), inflammation (C-reactive protein (CRP) > 5mg/L or α-1-acid glycoprotein (AGP) > 1g/L), or known malaria, were excluded.

Table 17.7. Two different sets of hemoglobin cutoffs used to define anemia. The INACG / WHO / UNICEF cutoffs are for individuals l and Dreyfuss (1998).

	Hemoglobin cutoff value (g/L)		
Age (y)	Males and Females	Males	Fema
NHANES III (from Looker et al., 1997).			
1-2	< 110	-	-
3-5	< 112	-	-
6-11	< 118	-	-
12-15	-	< 126	< 119
16-19	-	< 136	< 120
20-49	-	< 137	< 120
50-69	-	< 133	< 120
≥ 70	-	< 124	< 118

Age (y)	Hemoglobin cutoff value (g/L)		
	Males and Females	Males	Fema
INACG / WHO / UNICEF			
0.5-5	< 110	-	-
5-11	< 115	-	-
12-13	< 120	-	-
Men	-	< 130	-
Pregnant women		-	< 110
Nonpregnant women		-	< 120

The intersurvey variance around the Hb 5th percentile was low (3.5% for women and 3.6% for children) supporting the appropriateness of pooling multinational Hb data to derive a single-population group specific 5th percentile cutoff. The pooled 5th percentile estimates for apparently "healthy" individuals were: 96.5g/L (95% CI, 92.6-100.4g/L) for children aged 6-59mos, and 108.1g/L (95% CI, 103.5-112.7g/L) for nonpregnant women aged 15-49y, both lower than the WHO, 1968 estimates for these age groups shown in Table 17.7. Moreover, when the Hb fell below 96.1g/L for the children and below 110.1g/L for the women there was a compensatory increase in erythropoiesis, as indicated by a plot of Hb concentrations against sTfR (a biomarker of tissue iron deficiency and a physiological indicator of erythropoiesis) (Section 17.9). Nevertheless, the investigators highlight the need for more studies based on clinical outcomes to further confirm the validity of these revised Hb cutoffs for anemia for these two age groups; see Addo et al. (2021) for more details.

Cutoffs for infants and young children differ across countries and agencies. For iron-replete, breastfed Swedish infants aged 4-9mos, cutoffs defined by Domellöf et al. (2002) for anemia range from < 105g/L at 4mos; < 105g/L at 6mos; and < 100g/L at 9mos. In contrast, WHO (1968) set a single cutoff for infants and young children aged 0.5-5y of < 110g/L, higher than the pooled 5th percentile Hb cutoff reported above (96.5 g/L) (Addo et al., 2021). Further, Addo et al. report that infants aged 6-11mos had lower Hb levels than those of children > 48mos (P< 0.001) after adjustment for sex, Hb assessment method, and survey. Several other investigators have questioned the use of the same Hb cutoff for infants 6-11mos and young children (Emond et al., 1996; Sherriff et al., 1999; Wharton, 1999). The American Academy of Pediatrics recommends a cutoff of < 110g/L for children aged 1-3y (Baker et al., 2010), which is within the range set by Looker et al. (1997) for U.S. children (Table 17.7).

Cutoffs for adults are also shown in Table 17.7. For non-pregnant women, cutoffs set by WHO (1968) and Looker et al. (1997) are comparable, although higher than the multinational cutoff (108.1g/L) proposed for non-pregnant women (15-49y) by Addo et al. (2021). Sekhar et al. (2017) have also revised the cutoff-values for detecting iron deficiency anemia among non-pregnant women of reproductive age (120g/L) shown in Table 17.7. They used receiver operator characteristic (ROC) curves based on U.S. NHANES III data (2003-2010) to define Hb cutoffs for iron deficiency anemia with enhanced sensitivity, in an effort to improve the detection of iron deficiency in this population. These were: < 128g/L for women of reproductive age; < 127g/L for older women; < 125g/L for black women; < 130g/L for non-black women. Iron deficiency was defined by body iron calculated using serum ferritin and transferrin receptor (Section 17.10.1). Note that the selection criteria applied by Sekhar et al. (2017) to define a healthy U.S population were not as rigorous as those used by Addo et al. (2021) for their multinational study.

Table 17.8. Hemoglobin levels (g/L) to diagnose anemia at sea level. Source: Hemoglobin concentrations for the diagnosis of anemia severity. Geneva: World Health Organization (2011).

		Anemia — Hemoglobin (g/L)		
Population	**No anaemia**	**Mild**	**Moderate**	**Seve**
Children, 6-59mos	≥ 110	100-109	70-99	< 70
Children, 5-11y	≥ 115	110-114	80-109	< 80
Children, 12-14y	≥ 120	110-119	80-109	< 80
Non-pregnant women, ≥ 15y	≥120	110-119	80-109	< 80
Pregnant women	≥ 110	100-109	70-99	< 70
Men, ≥ 15y	≥ 130	110-129	80-109	< 80

Note that the Hb cutoff defined by Sekhar et al. (2017), based on U.S. NHANES data for non-pregnant women aged 12-49y, is higher (128g/L) than the pooled 5th percentile estimate based on non-pregnant women aged 15-49y in the multinational study (108.1g/L). Differences in the exclusion criteria applied, study settings, age range, and the statistical methods used to derive the cutoff values may account for some of this discrepancy, emphasizing the importance of careful selection and justification for any Hb cutoff applied.

Table 17.9. Classification of anemia as a problem of public health significance. Source: Hemoglobin concentrations for the diagnosis World Health Organization (2011).

Trigger levels for the prevalence of anemia (%)	Category of public health significance
≤ 4.9	No public health problem
5.0-19.9	Mild public health problem
20.0-39.9	Moderate public health problem
≥ 40	Severe public health problem

WHO have also adopted “thresholds” to classify anemia as mild, moderate, or severe (Table 17.8). These thresholds can be used in conjunction with the “trigger” levels based on the prevalence of anemia in each category to determine the public health significance of anemia in a population, as shown in Table 17.9.

Cutoffs for the elderly are also included in Table 17.7. Note that no specific Hb cutoff has been set by WHO for the elderly, even though there is some support for such an approach (Röhrig, 2016). As noted above, Sekhar et al. (2017) recommended a Hb cutoff of < 127g/L for older U.S. women based on ROC curves.

Cutoffs during pregnancy that differ according to trimester should now be used. WHO has adopted the CDC trimester-specific cutoffs to identify anemia in pregnant women (WHO, 2016); these are: 1st trimester: 110g/L; 2ndtrimester: 105g/L; 3rd trimester: 110 g/L. They were derived from gestational month-specific 5th percentile values for pooled data from four European studies with very small samples sizes.

Table 17.10. Thresholds for individuals according to deviations from the new normative trajectories for hemoglobin (Hb) in pregnancy

Individual women (for clinical use)	
Gestational age-specific cutoff (normative percentile)	**Probable diagnosis**
< 3rd percentile	Low Hb concentration
3rd-4.99th percentile	At high risk of low Hb concentrat
5th-9.99th percentile	At moderate risk of low Hb conce
≥ 10th percentile	Normal Hb

More recently, Ohuma et al. (2020) have developed gestational-age specific cutoffs derived from the percentiles presented in Appendix 17.2. Cutoffs to determine whether the Hb concentration for a pregnant women deviates from the gestational-age-specific normal Hb concentration (defined by Hb ≥ 10th percentile) are shown in Table 17.10. Figure 17.4 shows the summary estimates of the association of maternal Hb concentration (g/L) measured at any point during pregnancy and low birthweight by Hb cutoffs ranging from ≤ 70 to ≥ 140g/L. In this meta-analysis, low maternal Hb (at any time during pregnancy by cutoff) was associated with an increased odds of low birthweight (OR 95% CI): 1.42 (1.31-1.55). For high maternal Hb, the trend was similar (OR 95% CI): 1.80 (0.86-3.77) but not significant.

In some pregnancy studies, a fixed high Hb cutoff > 130g/L has been applied (Oaks et al., 2019), despite different gestational ages at the time of blood sampling. Uncertainties remain about the use of a fixed high Hb cutoff and the reason for the higher Hb concentrations; whether they are associated with failure to expand plasma volume, which in turn is associated with increased blood viscosity and decreased placental perfusion, is not clear (Fisher and Nemeth, 2017).

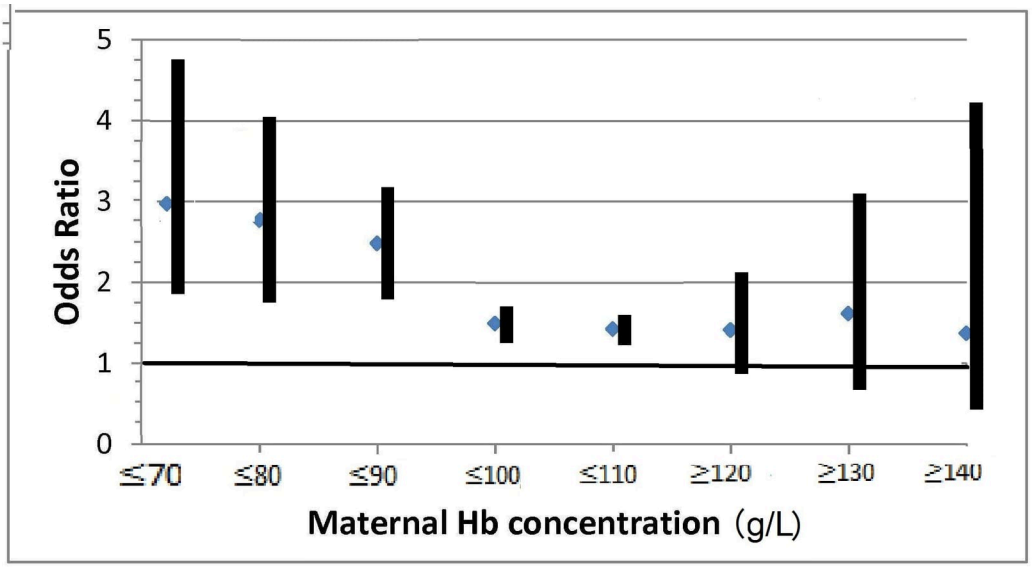


Figure 17.4. Meta-analysis summary of estimates of the association of maternal Hb concentration (g/L) measured at any point during pregnancy and low birthweight by Hb concentration cutoffs. Redrawn from Young et al. (2019).

17.2.4 Using hemoglobin distribution to assess population iron status

The distribution of Hb can be used in large-scale field studies to assess the iron status of a population (Yip et al., 1996). This simplified approach is designed to assess the prevalence and etiology of anemia, based solely on Hb, and is especially useful when it is not feasible to use multiple biochemical tests for iron status because of cost or operational constraints. It is also useful in developing countries where factors other than inadequate intakes of dietary iron, such as parasitic infections and genetic Hb

disorders, often affect red cell production and thus interfere with the interpretation of iron status measures.

The Hb distribution approach involves comparing the Hb distribution curves for men, women, and children of the study population, with optimal Hb distributions derived from a “healthy” reference sample.

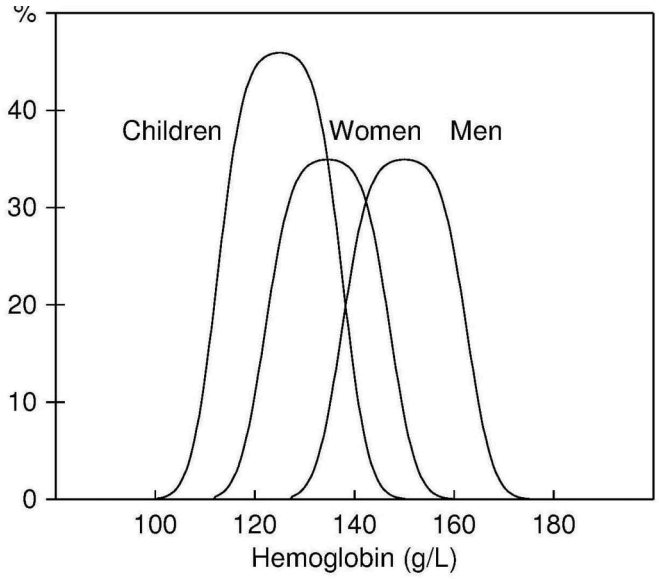


Figure 17.5 Distribution of hemoglobin in children aged 1-5y and in women and men aged 18-44y. Data from NHANES II after exclusison of subjects with abnormal values for indicators of iron status. Redrawn from Yip et al. (1996).

These have been compiled from U.S. NHANES II (Pilch and Senti, 1984) and NHANES III (Looker et al., 1997) populations by excluding subjects with biochemical evidence of iron deficiency, as noted earlier. Figure 17.5 depicts the Hb distributions for the NHANES II reference samples. These distributions can be used as a standard for comparison with the Hb distributions from other surveys.

If anemia is prevalent in the target population then the Hb distribution will be shifted to the left relative to the reference. The distribution approach can also indicate when inadequate dietary intake of iron is the main factor causing iron deficiency in a population. If this is the case, then the Hb distributions for children and women are both affected. Both subgroups will have significantly lower median Hb values when compared with their respective reference distributions, whereas the median of the distribution for adult men in the target population is virtually unaffected. For example, comparison of the Hb distributions for school-aged children from Zanzibar with the corresponding U.S. reference sample for African Americans shows a marked shift of Hb concentrations (Figure 17.6). In Zanzibar, the decreased Hb levels were related to gastrointestinal blood loss as a result of hookworm infection rather than inadequate dietary iron intake.

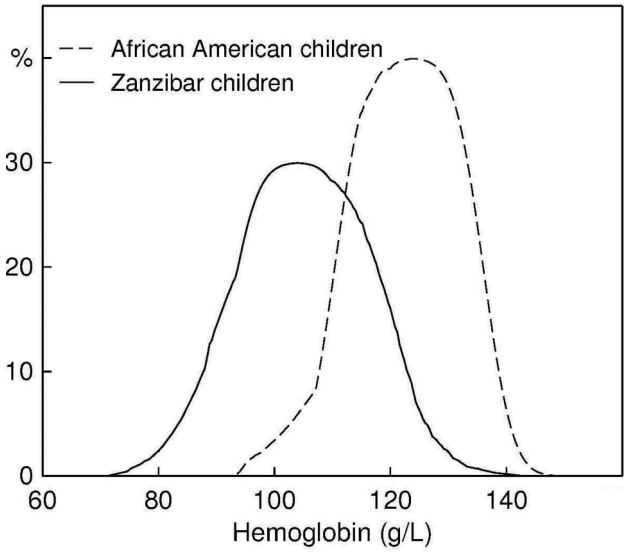


Figure 17.6. Distribution of Hb in a reference population of African American children with no biochemical signs of iron deficiency and school-age children from Zanzibar. Redrawn from Yip et al. (1996).

17.2.5: Measurement of hemoglobin

Hemoglobin can be determined in fasted or non-fasted blood samples and is best determined using venous blood, anticoagulated with EDTA. Alternatively, capillary blood from the heel, ear, or fingerpricks, collected in heparinized capillary tubes can be used, although less precise compared to measurements on venous blood. This is primarily because interstitial fluid may dilute capillary samples (Burger and Pierre-Louis, 2003).

The cyanmethemoglobin method, recommended by the International Committee for Standardization in Hematology (ICSH, 1987), is the most reliable, provided that the blood specimen has been accurately diluted. Incorrect dilution of the sample is one of the main sources of error in this method (Pilch and Senti, 1984). The method involves converting all of the usually encountered forms of Hb (oxyhemoglobin, methemoglobin, and carboxyhemoglobin) to cyanmet-hemoglobin. The analysis is then normally performed with a spectrophotometer. The reagents for this method are light-sensitive and poisonous. The coefficients of both analytical and biological variation for Hb by the cyanmethemoglobin method using venous blood are often less than 4% (Worwood, 1996). A WHO international standard is available to assess the accuracy of the assay (ICSH, 1987). Hemoglobin can also be determined in this way from blood spots collected in the field on filter paper discs; levels remain unchanged during storage for 1mo (Feraudi and Mejia, 1987).

Alternatively, Hb is best measured using a Coulter Counter using EDTA-anticoagulated blood. Most automated cell counter analyzers use the cyanmethemoglobin method. A portable hemoglobin photometer can be used for population assessment in remote field settings. The “HemoCue” is battery operated and uses a dry reagent (sodium azide) in a microcuvette for direct blood collection and measurement. In a new model (Hb-301), cuvettes can be stored at an extended temperature range (10-40°C) compared to earlier models (15-30°C) and the use of a control cuvette is no longer necessary. The accuracy and precision of Hb values based on the HemoCue are comparable to those obtained using the cyanmethemoglobin method, provided standardized procedures for the sample collection and analysis are followed (von Schenck et al., 1986). Details of standardization procedures to enhance the accuracy and reliability of Hb concentrations measured by the HemoCue are given in Burger and Pierre-Louis (Burger and Pierre-Louis, 2003). or in the U.S. CDC Nutrition Survey Toolkit.

This page titled 14.2: Hemoglobin (17.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.3: Hematocrit or packed cell volume

The hematocrit is defined as volume percentage or volume fraction of packed red blood cells (RBC) in blood, sometimes termed the “packed cell volume” (PCV). Variation with age and sex is shown in Figure 17.7,

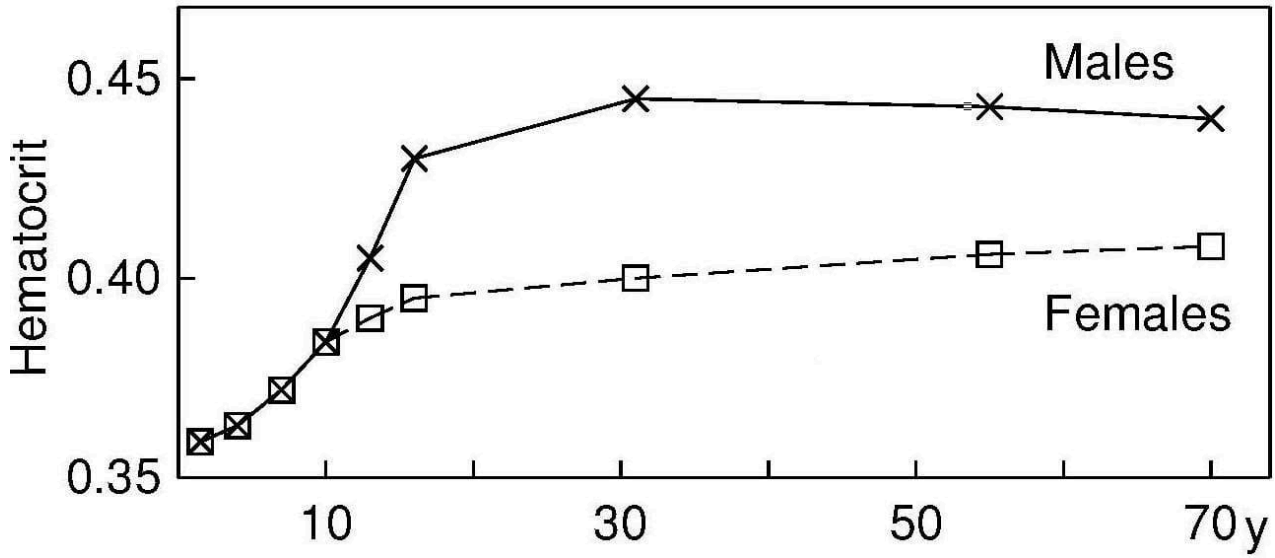


Figure 17.7. Median values for hematocrit by age and sex. Redrawn from Yip et al. (1984).

with higher concentrations in males than in females after puberty, as observed for Hb concentration (Zierk et al., 2011). As described for Hb, hematocrit concentrations decline steadily during pregnancy as a result of a greater increase in plasma volume relative to the increase in red blood cell mass, reaching a nadir at about 28-36wks in women not taking iron supplements. However, the hematocrit is only useful if Hb values are not available.

In iron deficiency, the hematocrit falls after Hb formation has become impaired. As a result, in early cases of moderate iron deficiency, a marginally low Hb value may be associated with a nearly normal hematocrit (Graitcer et al., 1981). Only in more severe iron-deficiency anemia are both Hb and hematocrit reduced. Limitations of the hematocrit determination are shown in Box 17.2.

Box 17.2: Limitations of the hematocrit determination

- Poor sensitivity as hematocrit falls only in the third stage of iron deficiency
- Poor specificity, as hematocrit is affected by the same factors that influence Hb
- Poor precision, particularly when capillary blood samples are used
- Values are age- and sex-dependent (Figure 17.7), with similar trends to those described for Hb.

Hematocrit, like Hb, is usually determined on EDTA-anticoagulated blood from a venipuncture or capillary blood. The measurement is relatively easy, rapid, and often used in screening for iron-deficiency anemia, although the technical errors for the measurement of hematocrit are greater than for Hb. Measurement errors may result from:

- Poorly packed iron-deficient cells, resulting in spuriously elevated values
- Improper mixing of blood caused by intermittent blood flow from the puncture site
- Excessive anticoagulant in the collection tube (e.g., if collection tube is not filled with blood)
- Dilution of blood by an alcohol solution because the finger was not dried before it was punctured
- Elevated white blood cell counts, leading to a poorly defined boundary between erythrocytes and plasma in the hematocrit

tube.

17.3.1 Interpretive criteria

Table 17.11. Hematocrit cutoffs used to define anemia. From Stoltzfus and Dreyfuss (1998).

Age (y) or group	Hematocrit cutoff (volume fraction)
0.5-5	< 0.33
5-11	< 0.34
12-13	< 0.36
Men	< 0.39
Nonpregnant women	< 0.36
Pregnant women	< 0.33

Results from two of the earlier U.K. Diet and Nutrition Surveys (Gregory et al., 1990; Gregory et al., 2000) present the mean, median, and lower and upper 2.5 or 5th percentiles for hematocrit for adults and older children respectively. Information dealing with U.K. National Survey years 9-11 (2016-2017 and 2018-2019) includes methodological details. A condensed ("zipped") comprehensive set of data tables based on information fromU.K. National Survey years 1-9 includes more recent data on Hb but not hematocit.

The hematocrit cutoff values at sea level used to define anemia and compiled by INACG / WHO / UNICEF, are presented in Table 17.11.

17.3.2 Measurement of hematocrit

Hematocrit can be measured manually by centrifuging a small amount of blood in a heparinized capillary tube until the red cells have been reduced to a constant packed cell volume. The hematocrit is calculated by comparing the height of the column of packed red cells with the height of the entire column of red cells and plasma, and expressed as the volume percentage or volume fraction of packed red blood cells. However, values measured in this way frequently have poor reproducibility, especially when the power supply is unstable and the equipment not well standardized (WHO, 2007).

Alternatively, hematocrit can be derived using automated electronic cell counters. These measure mean cell volume (MCV) and the number of red blood cells from which hematocrit is calculated as MCV × RBC concentration. Hematocrit values generated in this way tend to be about 1% higher than those generated manually (Looker et al., 1995).

This page titled 14.3: Hematocrit or packed cell volume is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.4: Red-cell indices

Red-cell indices are derived from measurements of Hb, hematocrit, and red-blood cell count. They can indicate cell size and the concentration of Hb within the cell allowing different types of anemia to be diagnosed. Fresh samples of whole anticoagulated blood are required. The accuracy, precision, and use of red-cell indices have increased with the growing availability of automated electronic cell-counting instruments.

Red-cell indices lack specificity. If subnormal values are noted in the absence of thalassemia trait, anemia of chronic disease, and other known causal conditions, additional measures of iron status are recommended to confirm the diagnosis of iron deficiency. The confirmatory tests commonly used are serum iron and total iron-binding capacity, serum ferritin, erythrocyte protoporphyrin, and more recently, soluble transferrin receptor (sTfR). These tests are discussed in Sections 17.6 to 17.9.

Table 17.12 summarizes the expected levels of the more important red-cell indices during iron-deficiency anemia, macrocytic anemia resulting from vitamin B_{12} or folic acid deficiency, and the anemia of chronic disease. The latter arises from inflammation due to infectious diseases (as noted earlier in “Acute and chronic inflammation” in Section 17.2.1), when pro-inflammatory cytokines stimulate the production of hepcidin. This in turn reduces both the absorption of dietary iron and the release of iron from body stores and from macrophages (which recycle iron from senescent erythrocytes). If the inflammatory condition persists, the supply of iron available for erythropoiesis is reduced, and anemia may develop (Anderson and Frazer, 2017). In chronic inflammation erythropoietin production may also be impaired and red-cell survival diminished (Raiten et al., 2015). The anemia of chronic disease is generally mild, and can be normocytic and normochromic (i.e., the average size and Hb content of the red blood cells are within normal limits) or microcytic and slightly hypochromic (i.e., red blood cells are small and contain less Hb than normal) (Andrews, 1999).

Table 17.12. Expected levels of the red-cell indices during iron-deficiency anemia, macrocytic anemia, and anemia of chronic disease Wintrobe et al. (1981).

Red cell index	Iron-deficiency anemia (microcytic hypochromic)	Macrocytic anemia (macrocytic)
MCV	Low	High
MCH	Low	High
MCHC	Low	Normal

Changes in red-cell indices during vitamin B_{12} and folic acid deficiency are described in more detail in Chapter 22. In iron-deficiency anemia, the fall in Hb is followed by a fall in mean cell volume (MCV), then mean cell Hb (MCH), and finally mean cell Hb concentration (MCHC). With the widespread use of automated hematology analyzers, MCV has become a key red cell parameter used to classify anemia (Table 17.12), although seldom used in the diagnosis of iron deficiency today.

Reference ranges derived from U.S. NHANES III stratified by nine age groups, sex and racial categories are available, based on MCV, MCH, MCHC, and red-cell distribution width (RDW). These are based on Coulter Counter data in which rigorous criteria were applied to exclude individuals with evidence of acute or chronic disease. Representative plots for each red-cell index by sex for non-Hispanic white; non-Hispanic black; Mexican Americans with the following percentiles are depicted: 2.5^{th}, 5th, 50th, 80th, 95th, 97.5^{th} (Cheng et al., 2004).

17.4.1 Mean cell volume

Mean cell volume or mean corpuscular volume (MCV) is a measure of the average size of the red blood cells expressed in femtoliters (fL). Cells may be abnormally large (macrocytosis), as in vitamin B_{12} or folic acid deficiency, or abnormally small (microcytosis), as in iron and vitamin B_6 deficiency. Mean cell volume is best determined directly with a Coulter Counter as results obtained in this way are highly reproducible. Reported within-subject biological variation for MCV is low, varying between 0.6 and 1.3% (Buttarello, 2016). If a Coulter Counter is not available, MCV can be calculated from the hematocrit and the red blood cell count determined manually (Wintrobe et al., 1981):

$$\text{MCV(fL)} = \frac{\text{hematocrit (volume fraction)}}{\text{red blood cell count } (10^{12}/\text{L})}$$

Mean cell volume is less affected by sampling errors in skin puncture capillary blood samples than Hb, because red-cell size is unaffected if the sample is diluted by the interstitial fluid.

A low MCV value only occurs when iron deficiency is severe. It is a relatively specific index for iron-deficiency anemia, provided that the anemia of chronic disease, certain genetic Hb disorders (α- and β-thalassemias including heterozygotes), and lead poisoning are excluded. In macrocytic anemias associated with vitamin B_{12} or folate deficiency, MCV values are high (Table 17.12). In pregnancy, MCV decreases slightly between 26 and 38wks gestation, likely attributed to the transfer of placental iron being most intense at this time, thus decreasing the iron availability for maternal erythropoiesis (Fisher and Nemeth, 2017). Spuriously high values for MCV may be apparent as a consequence of hyperglycemia or hypernatremia (high level of sodium in the blood).

The MCV increases progressively from 6mos to early adulthood (Yip et al., 1984). Differences according to sex are small; MCV is slightly higher in young adult females than in males. Individuals of African descent have lower MCV values than do Caucasians. Such differences are said to have a genetic basis, as noted for Hb under "Ethnicity" (Section 17.2.1).

Percentile distributions for MCV levels by race, sex, and age, for all persons 1-74y in the U.S. NHANES II population, are available, and for the U.S. NHANES II "healthy" reference population for all races as described for Hb in Section 17.2.2 (Pilch and Senti, 1984). Pregnant women and those with a higher risk of iron deficiency based on an abnormal value or missing value for 3 of the 4 tests (free erythrocyte protoporphrin, transferrin saturation, serum ferritin, and MCV) were excluded.

Cheng et al. (2004) report reference interval diagrams (2.5^{th}, 50th, and 97.5^{th} percentiles) based on the U.S. NHANES III database stratified by nine age groupings, sex, and race (i.e., Mexican American, Non-Hispanic Black, and Non-Hispanic White). Stringent exclusion criteria were applied to generate a "healthy" reference sample population (10-75y); details are available in Cheng et al. (2004).

Age and sex-specific data for MCV are also available for the older U.K. Diet and Nutrition surveys. Gregory et al., 1990; Gregory et al., 2000 present the mean, median, and lower and upper 2.5 or 5^{th}nbsp;percentiles for MCV for adults and older children respectively. Potentially iron-deficient individuals were not specifically excluded. Information dealing with U.K. National Survey years 9-11 (2016-2017 and 2018-2019) includes methodological details. A condensed ("zipped") comprehensive set of data tables based on information from the U.K. National Survey (years 1-9) includes more recent data on Hb and ferritin, but not MCV.

Table 17.13. Cutoffs for defining abnormal values of mean cell volume (MCV), transferrin saturation (TS), and serum ferritin (SF) that Nutrition Examination Survey (NHANES III). Subjects with high blood lead values were excluded. Iron deficiency was defined as hav

Age	Abnormal values	
	Mean cell volume (fL)	Transferrin saturation (%)
1-2y	< 74	< 10
3-4y	< 76	< 12
5y	< 77	< 12
15-49y	< 81	< 15

Cutoff values for MCV used to define abnormal values for the U.S. NHANES III survey were determined from values in the U.S. Healthy People 2000 Final Review; these are shown in Table 17.13 and differ slightly from those used earlier to identify abnormal values in the U.S. NHANES III phase 1 (1988-1991) (Dallman et al., 1996). Values for individuals with high blood lead values were excluded when defining these new cutoffs for three age groups for children plus those 15-49y. Values greater than 98fL indicate macrocytosis.

17.4.2 Mean cell hemoglobin

Mean cell hemoglobin (MCH) is the mean Hb content of individual red blood cells, and provides information that is comparable to that of MCV. It is derived from the ratio of Hb to the red blood cell count:

$$\mathrm{MCH(pg)} = \frac{\text{hemoglobin (g/L)}}{\text{red blood cell count } (10^{12}/\mathrm{L})}$$

Percentile distributions by age and sex for MCH are available for the older U.K. Diet and Nutrition surveys. Gregory et al., 1990; Gregory et al., 2000 present the mean, median, and lower and upper 2.5 or 5th percentiles for MCH for adults and older children respectively. Potentially iron-deficient individuals were not specifically excluded. Information dealing with U.K. National Survey years 9-11 (2016-2017 and 2018-2019) includes methodological details. A condensed ("zipped") comprehensive set of data tables based on information from the U.K. National Survey (years 1-9) includes more recent data on Hb and ferritin, but not MCH.

In contrast, stringent precautions were applied by Cheng and co-workers (2004) to generate a health-associated sample population based on the U.S. NHANES III survey from which composite reference intervals for MCH according to age and race, classified by sex were obtained; race-related trends in MCH were apparent. The MCH changed progressively from infancy to adulthood, when values ranged from 27-32pg. The MCH changes in iron-deficiency anemia are similar to those for MCV; MCH is low in iron-deficiency anemia but high in the macrocytic anemias of both vitamin B_{12} and folate deficiency (Table 17.12). In the latter, the red blood cells are laden with Hb but are reduced in number. In severe iron deficiency, the relative fall in MCH is greater than the corresponding fall in MCV (Dallman, 1977).

17.4.3 Mean cell hemoglobin concentration

If both the Hb concentration and the hematocrit are known, the concentration of Hb in the red blood cells can be determined. This is known as the mean cell hemoglobin concentration (MCHC) and is calculated as:

$$\mathrm{MCHC(g/L)} = \frac{\text{hemoglobin (g/L)}}{\text{hematocrit (vol. fraction)}}$$

The MCHC is normally determined with a Coulter Counter, although manual determinations are possible. After the first few months of life, MCHC is less affected by age than any other red-cell index (Matoth et al., 1971). Nevertheless, this index is the least useful of the red-cell indices because it is the last to fall during iron deficiency.

Percentile distributions by age and sex for MCHC are available for the older U.K. Diet and Nutrition surveys. Gregory et al., 1990; Gregory et al., 2000 present the mean, median, and lower and upper 2.5 or 5th percentiles for MCHC for adults and older children respectively. Potentially iron-deficient individuals were not specifically excluded. Information dealing with U.K. National Survey years 9-11 (2016-2017 and 2018-2019) includes methodological details. A condensed ("zipped") comprehensive set of data tables based on information from the U.K. National Survey (years 1-9) includes more recent data on Hb and ferritin, but not MCHC.

In the U.S. NHANES III sample of Cheng et al. (2004), stringent exclusion criteria allowed reference intervals for MCHC based on a "healthy" reference sample to be generated.

Mean cell hemoglobin concentrations are low in iron-deficiency anemia but normal in the macrocytic anemia of vitamin B_{12} and folic acid deficiency, and in the anemia of chronic disease (Table 17.12). Values in normal adults range from 320-360g/L. Values of < 300g/L indicate hypochromia and are associated with advanced iron deficiency.

This page titled 14.4: Red-cell indices is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.5: Red-cell distribution width (17.5)

RDW is a measure of the variation in red-cell size (i.e., anisocytosis). It is normally expressed as the percentage coefficient of variation of the mean cell volume:

$$\mathrm{RDW}(\%) = \frac{\mathrm{SD\ of\ MCV\ (fL)} \times 100\%}{\mathrm{MCV(fL)}}$$

The RDW can be determined routinely as part of a complete blood count on many Coulter Counters. Some counters output only the standard deviation as the measure of RDW.

RDW increases in iron-deficiency anemia and was included as a marker of iron status in the U.S. NHANES III survey (Looker et al., 1995), when a single cutoff for RDW of 14% was used. Based on their results on adults and children, Looker et al. (1995), however, suggested that different cutoffs are needed for children and adults. In the neonate, RDW is elevated because the relatively larger red blood cells containing fetal Hb are replaced by smaller cells containing Hb A (Lynch et al., 2018).

In the early 1980's Bessman et al. (1983) classified anemic disorders based on a combination of a low, normal, and high MCV and two categories of RDW — normal (< 15.1%) and high (indicating anisocytosis). However, in view of the numerous reported exceptions to this classification, together with differences in the reference intervals for the RDW depending on the analyzer manufacturer, this classification is seldom used today (Buttarello, 2016).

Clearly, a high RDW is not specific for iron deficiency. High values also occur in folate or vitamin B_{12} deficiency and when iron and folate, or iron and vitamin B_{12} deficiencies coexist. Also, in some genetic Hb disorders (e.g. in S β-thalassemia, sickle cell disease (SS), and sickle cell trait (SC), RDW is elevated, whereas in the anemia of chronic disease, RDW is normal. The interpretation of RDW is further complicated by the wide distribution of RDW values within a specific disease, and the discrepancy in cutoffs using Coulter Counters from different manufacturers (Buttarello, 2016).

Percentile distributions for RDW are available based on the U.K. and New Zealand national surveys. In the U.K. mean, median, and lower and upper 2.5 percentiles by age and sex are presented (Gregory et al., 1990; Gregory et al., 2000), whereas in the New Zealand Children's Nutrition Survey, data on RDW by age, sex, race or ethnicity, and the prevalence of elevated values (defined as >14%) were compiled (Parnell et al., 2003).

Cheng et al. (2004) have compiled reference interval plots (2.5^{th}, 50th, and 97.5^{th} percentiles) based on data from a “healthy” reference sample of the U.S. NHANES III data. These data for persons 10-75y were stratified into nine groups by age, sex and race. A tendency for RDW to increase with age was noted, suggesting that age-specific cutoffs may be required.

This page titled 14.5: Red-cell distribution width (17.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)

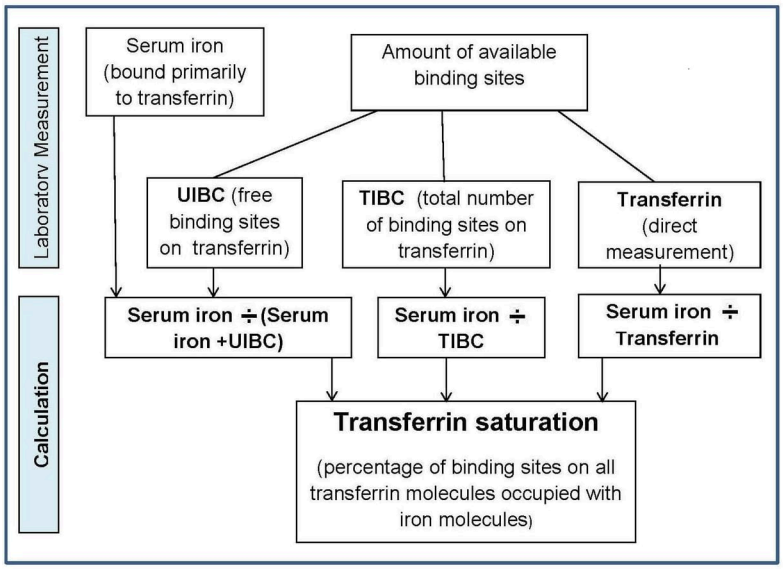


Figure 17.8. Laboratory measurements of iron indicators needed to calculate transferrin saturation. TIBC: total iron-binding capacity; UIBC: unsaturated iron-binding capacity. Modified from Pfeiffer and Looker (2017).

Three interrelated variables, serum iron, total iron-binding capacity (TIBC), and transferrin saturation have been used in the past for differentiating between nutritional iron deficiency and anemia arising from chronic infections, inflammation, or chronic neoplastic diseases; their inter-relationship is shown in Figure 17.8.

However, in view of the high biological variation of serum iron and transferrin saturation, they are used mainly in clinical settings to assess iron status. Today, in population surveys, serum iron and transferrin saturation have been replaced by serum ferritin and serum soluble transferrin receptor (sTfR).

Nutritional iron deficiency is characterized by an elevated serum transferrin and TIBC (not shown in Table 17.14), coupled with a low serum iron, and thus low transferrin saturation, whereas in chronic infections, inflammation, or chronic neoplastic diseases, concentrations of serum transferrin, serum iron, and TIBC are all low, so transferrin saturation is reduced (Table 17.14; Pfeiffer and Looker, 2017).

Table 17.14. Response of iron status indicators to a depletion of body iron compartments with and without concomitant inflammatior
Symbols: ↓ Decreased; ~ Normal; ↑ Increased; ACD, anemia of chronic disease; EP, erythrocyte protoporphyrin; IDA, iron deficiency a
chronic disease; NA, not applicable; S.Ferritin, serum ferritin; S.iron. serum iron; sTfR, soluble transferrin receptor; TSAT, transferrin s

Compartment	Indicator	IDA	ACD	IDA + ACD
Stored iron	S.Ferritin	↓	~ to ↑	↓ to ~
Transport iron	S.Iron	↓	↓	↓
	Transferrin	↑	↓	↓
	TSAT	↓	↓	↓
	EP	↑	↑	↑

Compartment	Indicator	IDA	ACD	IDA + ACD
	sTfR	↑	~	~ to ↑
Functional iron	Hemoglobin	↓	↓	↓
Inflammatory response	NA	~	↑	↑

Serum ironrepresents the fraction of iron that circulates bound to the iron transport protein, transferrin. The iron in the serum is derived mostly from iron recycled from catabolized red blood cells in the reticulo-endothelial system, with a concentration that normally ranges from8.9-21.5µmol/L(50-120µg/dL).Almost all the iron in the serum is bound to transferrin; non-transferrin bound iron (NTBI) usually comprises < 1% of the total iron pool in the serum (see Section 17.11.3 for details of NTBI).

In the fasting state, serum iron levels reflect the fraction of iron that circulates bound principally to transferrin and which is in transit from the reticulo-endothelial system to the bone marrow. After a meal, concentrations of serum iron rise due to the release of iron absorbed by enterocytes into the plasma pool. Hence, at this time, the measurement of serum iron will be representative of iron absorption. Serum iron is reduced by infection and inflammation.

Serum transferrin is the transport protein for serum iron. Theoretically each molecule of transferrin (MW ≈ 80kDa) is capable of binding to two moles of iron (MW ≈ 55.8kDa). When iron stores are exhausted, and serum iron concentrations are < 40-60µg/dL, serum transferrin increases in response to an increase in iron absorption. Hence, serum transferrin does not identify iron deficiency during the first stage of iron deficiency (i.e., iron depletion); for more details see Elsayed et al. (2016).

Total Iron Binding Capacity (TIBC) measures the total number of binding sites for iron atoms on transferrin. Hence, serum TIBC is closely related to, and often used as a proxy measure of, transferrin, rising once the iron stores are exhausted due to an increase in transferrin synthesis in response to increased iron absorption. Unlike serum iron, TIBC is not subject to rapid changes in concentration either within individuals or from day-to-day (i.e., circadian variation).

Transferrin saturation (as percent) is an estimate of the percentage of the two iron binding sites on all transferrin molecules that are occupied with iron. Transferrin saturation measures the iron supply to the erythroid bone marrow: as the iron supply decreases, the serum iron concentration falls and the saturation of transferrin decreases. Serum iron and TIBC are usually determined at the same time, so transferrin saturation is calculated as the ratio of serum iron to total iron binding capacity as shown below:

$$\text{Transferrin Sat. } (\%) = \frac{\text{Serum iron } (\mu\text{mol/L})}{\text{TIBC}(\mu\text{mol/L})} \times 100\%$$

Transferrin saturation can also be determined if unsaturated iron-binding capacity (UBIC) and serum iron are measured. In this case, TIBC is determined from the sum of serum iron and UIBC.

$$\text{TIBC}(\mu\text{mol/L}) = \text{UIBC}(\mu\text{mol/L}) + \text{ Serum iron } (\mu\text{mol/L})$$

Then

$$\text{Transferrin Sat. } (\%) = \frac{\text{Serum iron } (\mu\text{mol/L})}{\text{TIBC}(\mu\text{mol/L})} \times 100\%$$

Alternatively, transferrin saturation can be calculated from the ratio of serum iron to serum transferrin (Ritchie et al., 1999).

Transferrin saturation was one of the multiple iron biomarkers in the “ferritin model” employed in the U.S. NHANES II and U.S. NHANES III surveys (see Section 17.10). In healthy individuals, transferrin saturation (as percentage) usually ranges from 20-45%. Levels < 15% saturation indicate that the rate of delivery of iron to the developing erythrocytes is insufficient to maintain normal Hb synthesis, and hence are an early indicator of suboptimal iron supply. A prolonged period of time with transferrin saturation below 15%, results in iron deficient erythropoiesis which leads to changes in the number and shape of newly released reticulocytes and erythrocytes (WHO 2007). Nevertheless, low transferrin saturation levels are not specific for iron deficiency, as discussed below.

17.6.1 Factors affecting serum iron, TIBC, serum transferrin, and transferrin saturation

Biological variation for serum iron may be quite large, resulting in large daily variation in transferrin saturation: coefficients of variation among individuals for serum or plasma iron may exceed 30% — mainly as a result of variation in the release of iron from the reticuloendothelial system to the plasma. As a result, more than one sample is needed for a clinical evaluation. Table 17.4 compares the overall variability of serum iron with Hb and serum ferritin as indicated by within-subject, day-to-day coefficient of variation (as %) for healthy subjects. In contrast, serum transferrin concentrations are steadier.

In general, values for serum iron and thus transferrin saturation tend to be elevated in the morning (about 35%, ranging from 20-55%) but decrease in the afternoon and evening. As a result, measurements should be determined on fasting morning blood samples. In this way, the effects of both recent dietary intake and diurnal variation are minimized. Note that among night-shift workers, the pattern of diurnal variation may be reversed (Lynch et al., 2018).

TIBC is less subject to biological variation than is serum iron as noted earlier, but it is more susceptible to analytical errors. Day-to-day variation is about 8-12%

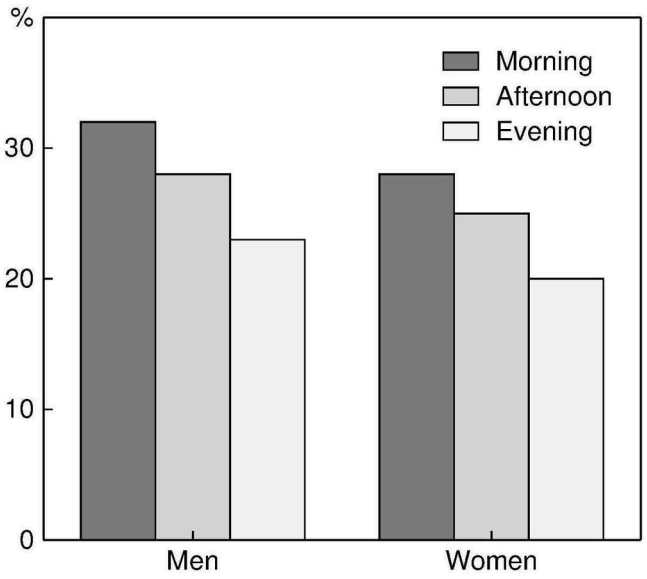


Figure 17.9. Median transferrin saturation (as %) by time of blood collection, Caucasian reference sample 20-49y, U.S. NHANES III phase 1. From Looker et al. (1995).

The large biological variation for serum iron has obvious implications for its use at both the population and individual levels. Furthermore, these limitations highlight the importance of using multiple indices of iron status to provide a more valid assessment of iron status than any single measurement (Section 17.10).

In U.S. NHANES III, the large diurnal variation for serum iron is reflected in the results for transferrin saturation shown in Figure 17.9. This measure was included in U.S. NHANES II and III in the earlier MCV and ferritin models used to assess iron status, but has now been abandoned.

Age-related changes in serum iron and serum transferrin are marked. Serum iron and transferrin levels rise during childhood until about the second decade of life (Figure 17.10). Values decline steadily with advancing age (Yip et al., 1984).

Sex has only a small effect on serum iron and transferrin saturation until the teenage years, after which levels in males increase whereas those for females remain relatively constant (Figure 17.10). In U.S. NHANES II, there was a tendency for serum iron levels to be higher in males than females after the middle of the second decade of life (Pilch and Senti, 1984). In U.S. NHANES III, the

median transferrin saturation was 26-30% for men and 21-24% for women (Appendix 17.3). This same trend of higher values in adult males than females 20-70y, was also seen in the median transferrin saturation values compiled by Ritchie et al. (2002).

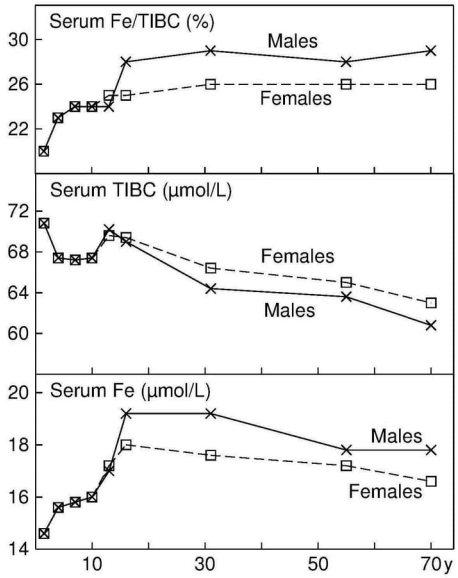


Figure 17.10. Median values for serum iron/total iron binding capacity (TIBC) (%), serum TIBC (µmol/L), and serum iron (µmol/L), by age and sex. Redrawn from (Yip et al., 1984).

Iron deficiency uncomplicated by the inflammatory response, leads first to a gradual reduction in iron stores, that once exhausted, results in a fall in serum iron levels. The low serum iron (and serum ferritin) levels down-regulate hepcidin synthesis, thus promoting iron absorption and greater iron release from storage sites into the circulation. As a consequence, serum transferrin increases to enhance the delivery of iron to the bone marrow to maintain effective erythropoiesis, so TIBC rises, resulting in a low transferrin saturation (Elsayed et al., 2016; Table 17.14).

Because these changes appear during the second stage of iron deficiency — termed iron deficient erythropoiesis — when iron stores are exhausted, transferrin saturation is a more sensitive index of iron status than the red-cell indices discussed earlier in Section 17.4, but less sensitive than serum ferritin (Section 17.7). Transferrin saturation is more consistently useful for diagnosing iron deficiency than either serum iron or total TIBC alone: a low transferrin saturation in association with an elevated TIBC is specific for iron deficiency. Nevertheless, because of the diurnal variation in levels, and the confounding effect of inflammation (see below), percentage transferrin saturation is not a suitable biomarker for population surveys.

Oral contraceptive agents and estrogen replacement therapy (i.e., conditions of estrogen excess) may lead to an increase in transferrin synthesis that is independent of body iron status. Levels of TIBC are also elevated, similar to those observed in iron deficiency (Pilch and Senti, 1984).

Pregnancy leads to a steady decline in serum iron as the plasma volume expands, whereas serum transferrin concentration increases, resulting in a decrease in transferrin saturation. Because of these rapid changes during pregnancy, serum iron and transferrin saturation are rarely used for diagnosing iron deficiency in pregnancy (Fisher and Nemeth, 2017). The magnitude of these changes is less in women supplemented with iron during pregnancy.

Iron overload results in an elevated body iron store and, thus, a high serum iron (and ferritin) level because of excessive absorption of iron and because there is no regulatory mechanism for controlling iron excretion. As a consequence, serum transferrin is reduced, TIBC falls, and transferrin saturation level is elevated (Table 17.14, Pfeiffer and Looker, 2017). In earlier U.S. NHANES III surveys, a transferrin saturation level > 70% was the criterion for detecting iron overload in adults. However, to detect iron overload today, WHO advises the inclusion of transferrin saturation and serum ferritin, plus details on known genetic and clinical risk factors (Garcia-Casal et al., 2018).

The most common genetic risk factor for iron overload is hereditary hemochromatosis, with a worldwide prevalence in Caucasian

populations of 3.5-4.5per 1000 population. It is more common in males. Clinical risk factors for othertypes of iron overload include hemolytic anemia, chronic liver disease, long-term transfusion therapy for genetic Hb disorders, and African iron overload (Elsayed et al., 2016).

Anemia of chronic disease is associated with infectious diseases (see Section 17.2.1) that induce the acute-phase response, the local reaction of which is inflammation. Pro-inflammatory cytokines promote the release of hepatic hepcidin which in turn blocks the release of iron from the body stores into the circulation (independent of iron status) and inhibits absorption of dietary iron. Therefore, iron is no longer available for transport to the bone marrow for erythropoiesis. Consequently, concentrations of both serum iron and serum transferrin fall, so levels of TIBC are low. Hence, percentage transferrin saturation is decreased, a response primarily reflecting the decreased serum iron levels. Therefore, serum transferrin acts as a negative acute phase protein, decreasing in the anemia of chronic disease, whereas in iron deficiency anemia, serum transferrin concentrations are increased (Table 17.14).

The rate at which serum iron concentrations return to normal after the acute phase response varies among individuals, sometimes normalizing with 24-48hrs of an acute infection. In contrast, during chronic inflammatory states such as arthritis, serum iron levels may remain low for a prolonged period of time, thus decreasing the availability of iron to cells leading to the anemia of chronic disease. Such variability in the response causes uncertainty in the interpretation of serum iron in inflammation (WHO, 2007).

Obesity is associated with lower serum iron and tansferrin concentrations leading to iron deficiency unrelated to dietary intakes; Hb concentrations are not reduced (Menzie et al., 2008).

Table 17.15. Hematologic characteristics, iron status, and inflammatory biomarkers in normal weight (NW) and overweight/obese (OV and OW/OB was assessed with independent samples t-test. P-values < 0.05 were considered significant. NW (BMI 18.5-24.9kg/m^2), total iron binding capacity, sTfR soluble transferrin receptor, IQR interquartile range. Values are: [a] mean (± s.d.), and [b] geometric me from OW/OB. Data from Cepeda-Lopez et al. (2019).

	NW n=24
Blood volume (mL/kg)[a]	74.9 ± 5.8[c]
Blood volume (L)[a]	4.59 ± 0.46[c]
Plasma volume (mL/kg)[a]	46.6 ± 4.8[c]
Plasma volume (L)[a]	2.86 ± 0.30[c]
RBC (mL/kg)[a]	28.2 ± 2.1[c]
RBC volume (L)[a]	1.74 ± 0.22[c]
Hb (g/dL)[a]	13.5 ± 0.9
Hb mass (g/kg)[a]	9.2 ± 0.7[c]
Total Hb (g)[a]	566 ± 71[c]
Serum iron (µg/mL)[a]	1.06 ± 0.44
Total serum iron (mg)[a]	3.02 ± 1.28
TIBC (µg/mL)[a]	3.68 ± 0.66
Total TIBC (mg)[a]	10.5 ± 2.28
Transferrin saturation (%)	29.8 ± 13.4
sTfR (mg/L)[b]	6.61 (5.86, 7.45)

	NW n=24
Total sTfR (mg)[b]	18.8 (15.0, 21.5)[c]
SF (µg/L)[b]	50.6 (40.2, 63.4)
Total SF (mg)[b]	0.14 (0.10, 0.18)[c]
Body iron (mg/kg)[a]	5.64 ± 2.70
Serum hepcidin (ng/mL)[a]	9.20 ± 6.44
Total hepcidin (µg)[a]	25.5 ± 17.6
IL6 (pg/mL)[b]	0.69 (0.53, 0.91)[c]
Total IL6 (ng)[b]	1.95 (1.36, 2.63)[c]
CRP (mg/dL)[b]	1.05 (0.60, 1.85)[c]
Total CRP (mg)[b]	29.87 (15.4, 53.5)[c]
AGP (g/L)[a]	0.79 ± 0.2[c]
Total AGP (g)[a]	2.24 ± 0.5[c]

The mechanisms involved may include hemodilution arising from an increased plasma volume in overweight and obese individuals and/or the presence of adiposity-related inflammation. Support for these two mechanisms has been provided by an elegant study in which plasma volume and inflammation were measured (Cepeda-Lopez et al., 2019).

Data in Table 17.15 confirm both a higher plasma volume and the presence of adiposity-related inflammation in the overweight / obese (OW/OB) women (BMI 25-39.9kg/m^2) compared to their normal-weight peers (BMI 18.5-24.9kg/m^2). Note the elevated serum concentrations and total mass of IL-6 levels in the OW/OB women. These higher levels stimulate an increase in circulating hepcidin which, in turn, is likely to impair iron absorption, and induce the hypoferremia (i.e., low serum iron levels) observed in the OW/OB women compared to their normal weight peers (Cepeda-Lopez et al., 2019). There is also an increase in Hb mass (i.e., total Hb in g) in OW/OB women, thus increasing their iron requirements for erythropoiesis and contributing to the increased risk for iron deficiency in OW/OB women.

Decreased erythropoiesis reduces the demand for iron for Hb synthesis, so levels of serum iron remain at normal or slightly above normal levels, and TIBC is normal or low. Consequently, transferrin saturation may be high. This condition occurs with chronic renal failure, and aplastic anemia as well as with the action of certain drugs and toxins.

Increased erythropoiesis occurs in conditions in which there is an increased demand for iron for Hb synthesis. Consequently, serum iron concentrations fall below normal limits, whereas TIBC is often high-normal or even high, and levels of circulating serum transferrin increase. These same trends occur in iron deficiency, so additional measures of iron status (e.g., serum ferritin) should be used to distinguish between increased erythropoiesis and iron deficiency. Increased erythropoiesis occurs during the stages of the life cycle when needs for iron are increased, such as growth and pregnancy, after recovery from bone marrow depression, in response to vitamin B_{12} and folate therapy, hypoxia, and sometimes in polycythemia.

Hemolysis leads to falsely elevated serum iron concentrations because the concentration of iron in red blood cells is higher than that of serum. Therefore, hemolyzed serum samples should not be used for the measurement of serum iron.

17.6.2 Interpretive criteria — use of the distribution of reference values for serum iron, serum transferrin, TIBC, and transferrin saturation (as %)

Data shown below for the U.S., U.K., and New Zealand are based on an apparently “healthy” population sampled during a national nutrition survey. Also included is an additional set of U.S. data from a “healthy” reference sample group.

- U.S. Tables of mean and selected percentiles by age and sex, for percentage of transferrin saturation, derived from all the persons in the U.S. NHANES III survey, with the exception of breastfeeding infants and children, are given in Appendix 17.3. (IOM, 2001). A diagrammatic summary of this information by age and sex is shown in Figure 17.11.

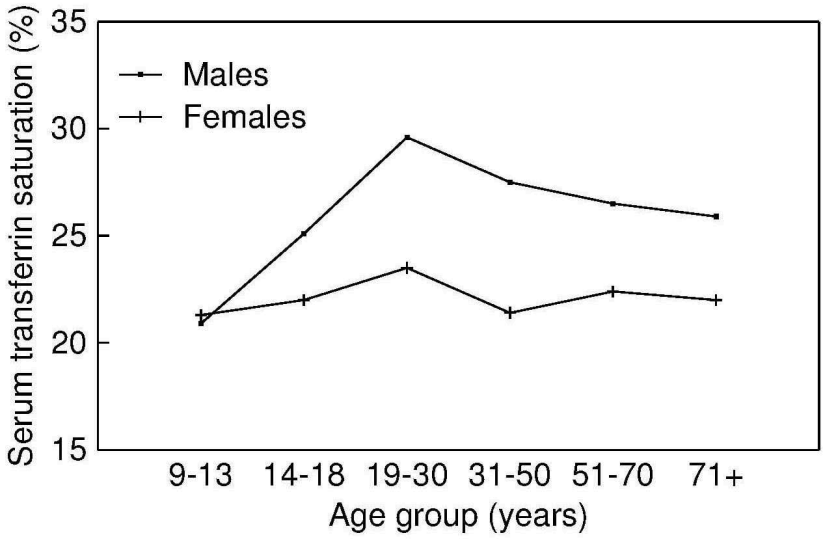


Figure 17.11. U.S. NHANES III median serum transferrin saturation by sex/age category. Data from IOM (2001).

- U.K. data from the older U.K. Diet and Nutrition surveys are available. Gregory et al., 1990 and Gregory et al., 2000 present the mean, median, and lower and upper 2.5 or 5th percentiles for percentage transferrin saturation, plasma iron, and TIBC for adults and older children respectively. Potentially iron-deficient individuals were not specifically excluded.
- New Zealand Tables of the mean percentage transferrin saturation by age, sex, race or ethnicity for school-aged children (5-15y), and the prevalence of low values (i.e., < 12% if ≤ 5y and < 14% if 6-15y) derived from the New Zealand National Children's Nutrition Survey(Parnell et al., 2003) are available.
- U.S. Tables of the median, 2.5th and 97.5th percentiles by age and sex for percentage transferrin saturation, serum iron, and serum transferrin derived from a “healthy” reference sample of U.S. Caucasians 1-80y are available. Individuals with evidence of inflammation (i.e., CRP > 10mg/L), suggestive evidence of iron excess (i.e., transferrin saturation levels > 80% for males and > 70% for females), and low serum iron values (i.e., < 5µmol/L) were excluded. Levels varied by age, generally rising until the second or third decade of life and then decreasing (Ritchie et al., 1999, 2002).

17.6.3 Cutoffs for serum iron, transferrin and transferrin saturation (as %)

Cutoffs for infants are not clearly defined because age-related changes in serum iron and transferrin are marked as shown in Figure 17.10. In newborns, serum iron concentrations vary markedly, prohibiting the use of percentage transferrin saturation as an indicator of iron status in neonates(Lynch et al., 2018).

Table 17.16. Cutoff points used for identifying abnormal values of iron status measures in the analysis of the NHANES III data. [a] *80µ*

Age group (y)	Transferrin saturation (%)	Serum ferritin (µg/L)
1-2	< 10	< 10
3-5	< 12	< 10
6-11	< 14	< 12
12-15	< 14	< 12

Age group (y)	Transferrin saturation (%)	Serum ferritin (µg/L)
≥ 16	< 15	< 12

Cutoffs for children for transferrin saturation (as percent) are given in Table 17.16, where they are classified into four age groups with values ranging from < 10% to < 14%. These cutoffs approximate the 12th percentile and are slightly lower than used in U.S. NHANES II because of a change in the serum iron method in U.S. NHANES III (see below) and the fact that more blood samples were drawn late in the day, when serum iron values are normally lower (Looker et al., 1997).

Cutoffs for adults were also compiled by Looker et al. (1997) with a transferrin saturation below 15% said to be associated with iron-deficient erythropoiesis, if subjects with infection and inflammation are excluded(Lynch et al., 2018).

Cutoffs for iron overload in adult males and females are based on transferrin saturation values > 70%. This value was used in both the NHANES II and III surveys, in combination with elevated serum ferritin levels (Section 17.7.1). In some studies, a lower cutoff value for transferrin saturation for screening for iron overload in women (i.e., > 50%) has been used (Hallberg, 1995).

17.6.4 Measurement of serum iron, TIBC, and serum transferrin

Serum iron can be assayed by a variety of methods, some of which can be performed using a clinical chemistry autoanalyzer. Colorimetric procedures frequently use ferrozine as the chromogen to react with Fe^{++} to form a violet complex (Giovanniello et al., 1968). Other chromogens that can be used include bathophenanthroline, sulfonate, and ferrozine (Elsayed et al., 2016). Alternatively, serum iron can be assayed directly by atomic absorption spectrophotometry. Often there are large discrepancies between serum iron concentrations assayed by different methods (Tietz et al., 1996), particularly at the lower end of the range where the diagnosis of iron deficiency is important (Worwood, 1996). Plasma iron concentrations normally range from 8.9-21.5µmol/L (50-120µg/dL), with those indicative of iron deficiency from 8.9-10.7µmol/L (50-60µg/dL). Factors that must be addressed when assaying for serum iron to ensure reliable results are shown in Box 17.3.

Box 17.3: Procedures to ensure reliable serum iron results

- Care must be taken to prevent adventitious contamination during collection and analysis of the serum/plasma samples
- Care is required to avoid hemolyzed samples which may falsely elevate iron concentrations
- Standardized protocols must be used for blood sampling to minimize variations due to time of day and post-prandial effects
- Well trained laboratory personnel to ensure accurate and reproducible results.

Modified from WHO (2007).

Total iron binding capacity can be calculated using a direct assay method that involves saturating the transferrin iron-binding capacity with excess ferric (Fe^{3+}) iron, and then removing the excess unbound iron by using an adsorbent agent such as magnesium carbonate, charcoal, or anion exchange resin. The iron that is fully saturating transferrin (i.e., transferrin-bound iron) in solution is then measured (ICSH, 1978).

Uncertainties in this assay include variations in the type and amount of magnesium carbonate used to remove unbound iron, and in the concentration of the saturating-iron solution (Pilch and Senti, 1984). In 1990, the ICSH published revised recommendations for this assay (ICSH, 1990).

An indirect method can also be used to calculate TIBC. In this method, serum is incubated at near neutral pH with a predetermined amount of Fe^{3+} that saturates all the available free sites on serum transferrin. A chromogen is then added to complex with the free unbound iron, and as the total amount of iron added is known, UIBC can be measured. TIBC is then calculated based on the sum of UIBC and serum iron as shown in Section 17.6.

Potential sources of error in the measurement of both serum iron and TIBC include contamination by exogenous iron, interference by lipids and bilirubin in the spectrometric methods, and a copper interference with colorimetric assays (Tietz et al., 1996). To overcome the copper interference, the assay employing ferrozine, used for serum iron and TIBC in U.S. NHANES III, was improved by the addition of 1% thiourea (Looker et al., 1997). This led to lower serum iron values (by about 1.6µmol/L) in this

survey than those for U.S. NHANES II, with a much larger discrepancy at lower concentrations than at higher ones (Looker et al., 1995). Quality-control sera, with certified values for serum iron and TIBC and covering a wide range, should be included with each assay. A reference material for serum iron is available from the National Institute for Standards and Technology (NIST SRM 937), although no reference preparation exists for TIBC. Iron-free water should always be used.

Serum transferrin can be assayed using several immunological assays involving immunochemical turbidimetry, nephelometry, radial immunodiffusion methods, and enzyme-linked immunosorbent assays (ELISA). Gambino et al. (1997) recommend using an automated nephelometric method to assay serum transferrin and not the manual radial immunodiffusion method. Variations between these different methods have been reported, some involving failure of the methods to take into account other iron-binding proteins (Elsayed et al., 2016). A certified reference material (CRM 470 reference preparation for protein in human serum RPPHS) is now available for serum transferrin.

Measurements of iron transport are now often omitted from nutrition surveys because of the large volume of serum sample required, cumbersome analytical methods, and poor sensitivity and specificity for iron deficiency; their greatest value is in distinguishing between the anemia of chronic disease and that of true iron deficiency (Table 17.14) (Cook, 1999).

This page titled 14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.7: Serum ferritin (17.7)

Ferritin was first identified in human serum by Addison et al. (1972), where it represents only a small fraction of the body's ferritin pool. Its function in the serum is unknown. Ferritin appears to enter the serum by secretions from hepatocytes or specialized macrophages (Cook and Skikne, 1982). The ferritin molecule consists of an intracellular hollow protein shell composed of 24 subunits that surrounds an iron core containing 4000-4500 iron atoms (WHO, 2020).

In most healthy individuals with no evidence of concurrent infection or inflammation, the concentration of serum or plasma ferritin parallels the total amount of storage iron (Cook et al., 1974). Once iron stores become exhausted, however, serum ferritin concentrations no longer reflect the severity of the iron-deficiency state. Serum ferritin is the only measure of iron status that can reflect a deficient, excess, or normal iron status. Evidence for the quantitative relationship between serum ferritin and storage iron is shown in Box 17.4.

Box 17.4: Evidence for the quantitative relationship between serum ferritin and storage iron

- Stainable bone marrow stores correlate positively with serum ferritin levels.
- Iron removal by phlebotomy produces a fall in serum ferritin levels, the decline paralleling changes in liver iron stores as measured by liver biopsy.
- Iron therapy and repeated blood transfusions generate a positive response in serum ferritin levels.

Translation of serum ferritin values into the size of the iron stores should be made cautiously, in view of the limited knowledge of this relationship and the confounding factors that influence serum ferritin concentrations. There is also considerable between-subject variation in the relationship between serum ferritin and tissue iron. In otherwise healthy individuals, a serum ferritin concentration of 1µg/L is said to be equivalent to approximately 8-10mgof storage iron in an adult (Cook, 1999). During iron overload, however, the proportion of iron stored as insoluble hemosiderin increases. Besides providing a body store of iron, the uptake of iron into ferritin limits the capacity of iron to generate free radicals and prevents oxidative damage to tissues (Ghio et al., 2006).

17.7.1 Factors affecting serum ferritin

Biological variation in serum ferritin levels is less marked than for serum iron. The overall within-subject day-to-day coefficient of variation for serum ferritin in healthy subjects over a period of weeks is about 15% (see Table 17.4), compared with approximately 30% for serum iron: the diurnal variation appears to be minimal. Analytical variation across differing automated methods for serum ferritin ranges from a CV of 10-15%, whereas when the same analytical method is used the CV is < 10% (Blackmore et al., 2008).

Table 17.17. Changes in serum ferritin concentrations with age. The U.S. NHANES III results are abstracted from IOM (2001) and are the median values and representative of all races. Breast-feeding infants and children were excluded.

Age group (y)	Males (µg/L)	Females (µg/L)
1-3	22.9	
4-8	28.7	
9-13	32.9	30.6
14-18	46.7	27.0
19-30	112.5	36.0

Age group (y)	Males (µg/L)	Females (µg/L)
31-50	155.9	39.9
51-70	154.5	89.4
≥ 71	137.1	95.3
Pregnant	-	28.0
Lactating	-	32.9

Age-related changes in serum ferritin concentrations are relatively marked and differ according to sex. At birth, concentrations are high, as iron stores in the liver are abundant. During the first 2mo, serum ferritin levels rise further, as iron is released from the fetal red cells and the rate of erythropoiesis is slow. Serum ferritin concentrations then fall throughout later infancy (Domellöf et al., 2002a). At about 1y, ferritin values increase again, and continue to rise into adulthood (WHO, 2020, Table 17.17).

Sex differences in serum ferritin concentrations exist. Among infants, males have lower values than females at 4, 6, and 9mo, and these values are not responsive to iron supplementation (Domellöf et al., 2002a). As a result, there may be a need to develop sex-specific cutoffs for serum ferritin in infancy. During adolescence, females have lower values than males, a trend that persists into late adulthood, reflecting their lower iron stores due to menstruation and childbirth. Levels in males reach a maximum between 30 and 39y, after which levels remain constant until about 70y. In females, serum ferritin levels remain relatively low until menopause, after which they rise steeply (Table 17.17) (Zacharski et al., 2000; WHO, 2020).

Race is known to influence serum ferritin concentrations. In U.S. NHANES III, adult male African Americans had higher serum ferritin values than Caucasians and Hispanics of comparable age for each decade of life (Figure 17.12).

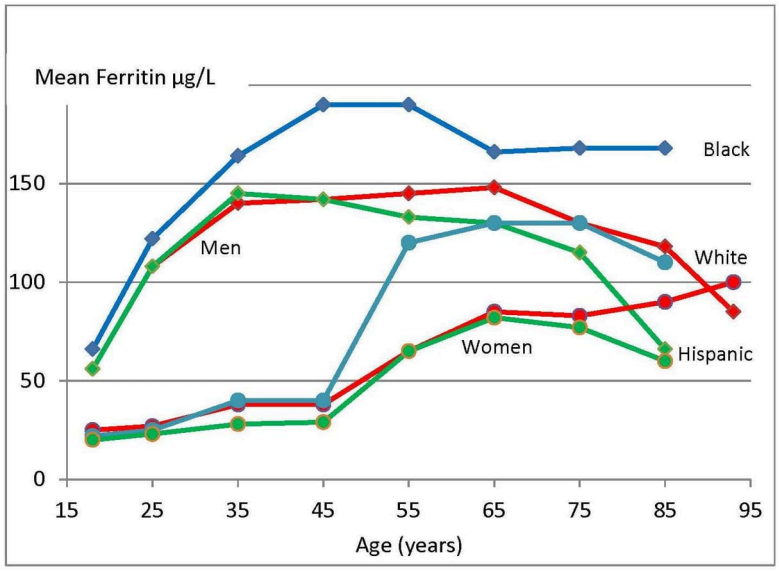


Figure 17.12. Distribution of mean serum ferritin concentration by age group for American men and women from the NHANES III survey by age and racial group. Redrawn from Zacharski et al. (2000).

In the adult females, serum ferritin levels were only higher in the African Americans after menopause. Overall, serum ferritin values (as geometric means) were approximately 7%-8% greater for the African Americans than for the Caucasians throughout the second half of life. Such differences are unlikely to be associated with increased intakes of dietary iron but, instead, are likely due to genetic factors (Zacharski et al., 2000).

Iron deficiency leads to lower iron stores: serum ferritin values fall progressively as the stores decline, and before the

characteristic changes in serum iron and TIBC. The decline in serum ferritin arises because hepcidin levels are suppressed in iron deficiency, thus facilitating the greater mobilization of iron from body stores. However, once body iron stores are depleted (i.e., serum ferritin concentrations < 15µg/L), serum ferritin concentrations are not indicative of the severity of iron deficiency (Pfeiffer and Looker, 2017). In frank iron-deficiency anemia, when classic microcytic hypochromic anemia occurs, serum ferritin levels are very low or zero. A low concentration of serum ferritin is characteristic only of iron deficiency (Dallman et al., 1980).

The decline in iron stores in uncomplicated iron deficiency has been confirmed in a systematic review by Garcia-Casal et al. (2018). Serum ferritin and stained bone marrow aspirates were measured simultaneously and compared in studies of otherwise healthy individuals without inflammation and with both normal iron stores and iron deficiency. In Table 17.18 the mean serum ferritin level for those with iron deficiency (based on the gold standard test of a bone marrow Fe content = 0µg/L) was lower (15.1µg/L (range of = 12-18µg/L)) compared to the mean serum ferritin value when iron was present in the bone marrow (i.e., bone marrow Fe content > 1µg/L). The mean difference in serum ferritin concentration between iron deplete and iron replete adults was approximately 55µg/L, as shown in Table 17.18 (Garcia-Casal et al., 2018).

Table 17.18. Mean ferritin concentrations (µg/L) in healthy adults classified as iron deficient or iron sufficient by bone marrow iron co systematic review. From Garcia-Casals et al. (2018).

	Apparently healthy adults	Non-healthy adults
Bone marrow Fe content = 0µg/L	15.10	82.43
Number of studies included	9 (n=390)	38 (n=1023)
Bone marrow Fe content ≥ 1µg/L	70.37	381.63
Number of studies included	3 (n=151)	38 (n=1549)

Acute and chronic inflammation increase levels of pro-inflammatory cytokines in the systemic circulation, which in turn elevate serum ferritin, an acute phase protein. At the same time, hepatic synthesis of hepcidin, also an acute phase protein, is increased, thereby inhibiting both the release of iron from body stores and the absorption of dietary iron from the gastrointestinal tract. Reductions in food intake, and hence iron intake during inflammation, may also impair iron status. Hence, the assessment of iron status in settings where concomitant infection and inflammation occur is challenging and may result in an underestimate of the prevalence of iron deficiency in a population (Suchdev et al., 2017). This is illustrated by the high mean serum ferritin concentration (i.e., 82.43µg/L) (n=1023) in Table 17.18 for non-healthy individuals, even though iron depletion had been confirmed by a bone marrow content = 0µg/L in these individuals (Garcia-Casal et al., 2018). For more details of the impact of inflammation on iron status indicators, the reader is referred to Raiten et al. (2015) and Lynch et al. (2018).

Elevated concentrations of biomarkers of acute and chronic inflammation are associated with misleadingly high serum ferritin concentrations during inflammation. The WHO (2020) recommends the concurrent measurement of serum C-reactive protein (CRP) and α-1-acid-glycoprotein (AGP), each providing a measure of the severity and duration of acute and chronic inflammation, respectively. The assay of CRP and AGP clearly aids the interpretation of serum ferritin concentrations in regions where infection or inflammation are widespread; see Section 17.7.3 for details of their use.

Pregnancy results in a gradual decrease in serum ferritin concentrations, reaching the lowest concentration in the third trimester, a decline partly attributed to hemodilution together with progressive depletion of iron stores in unsupplemented women. At this stage, if iron stores are exhausted, serum ferritin will no longer reflect the severity of the iron deficient state. In women supplemented with iron, however, the decline in serum ferritin in the third trimester is reduced (Figure 17.13).

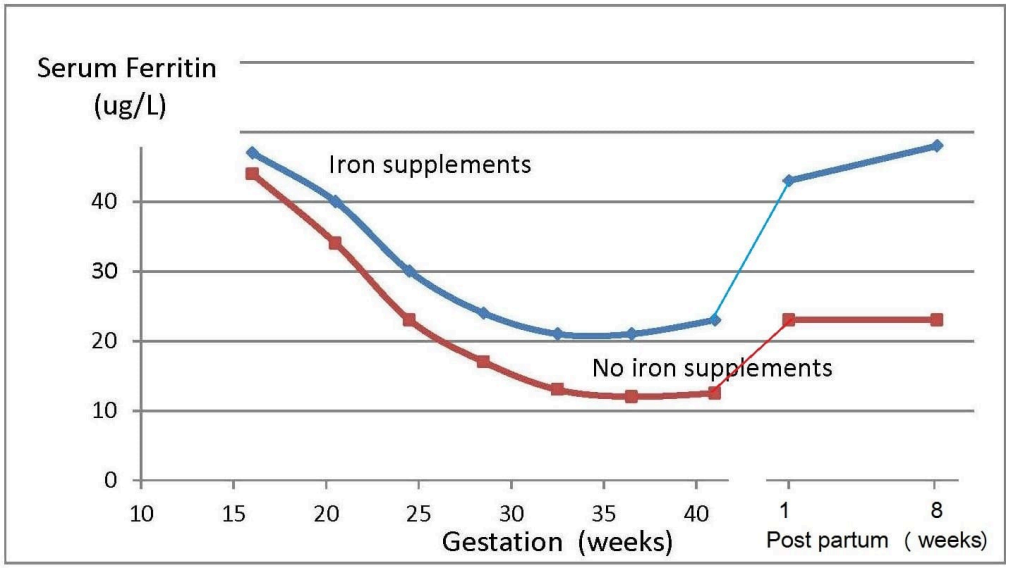


Figure 17.13. Geometric mean serum ferritin concentrations during pregnancy in 63 women with iron supplementation and 57 women without iron supplementation. Redrawn from Fisher and Nameth (2017).

In addition, because pregnancy is an inflammatory state, the potential effect of inflammation on serum ferritin concentrations must be taken into account in their interpretation. As a consequence, including concurrent measurement of inflammatory markers such as CRP and AGP with serum ferritin during pregnancy is essential.

In view of these uncertainties, the use of low serum ferritin concentrations alone to define iron deficiency during pregnancy is not recommended. Additional iron biomarkers such as soluble transferrin receptor (sTfR) should also be included (see Section 17.9.1) as used in the U.S. NHANES 1999-2006 survey (Mei et al., 2011).

Concerns over high serum ferritin concentrations during the second or third trimester have been raised in view of their association with adverse pregnancy outcomes such as the increased risk of preterm delivery (Scholl, 1998). Whether these high ferritin concentrations reflect excessively high maternal iron stores, the presence of inflammation in complicated pregnancies, or the failure of the plasma volume to expand, is unclear (Fisher and Nemeth, 2017).

Oral contraceptive agents have been associated with higher serum ferritin concentrations (Miller, 2014). Mechanisms proposed include lighter menstrual blood loss and possibly a direct relationship between estrogen and iron homeostasis. High levels of estrogen are said to inhibit the expression of hepcidin, thus increasing uptake of iron into the body (Yang et al., 2012).

Iron overload is associated with a high serum ferritin concentration arising from excessive absorption of iron. It is usually caused by autosomal recessive genetic conditions such as hereditary hemochromatosis, as well as other conditions shown in Table 17.19.

Table 17.19. Mean serum ferritin concentrations (µg/L) (n=number of individuals) with pathologies of different origins classified as iron overloaded or non-iron overloaded according to liver iron content determined by biopsies in studies included and extracted for the diagnostic test accuracy systematic review. From Garcia-Casals et al. (2018).

Disease/Condition	No iron overload (Liver iron concentration < 3.2mg/g dry liver wt. or defined by trialist)	Iron overload present (Liver iron concentration > 3.2mg/g dry liver wt. or defined by trialist)
Hemochromatosis	401.16 (n=78)	1043.82 (n=247)
Alcoholism, liver cirrhosis, liver disease	755.2 (n=154)	1322.63 (n=131)
Chronic hepatitis C	135.9 (n=526)	350.3 (n=213)

Disease/Condition	No iron overload (Liver iron concentration < 3.2mg/g dry liver wt. or defined by trialist)	Iron overload present (Liver iron concentration > 3.2mg/g dry liver wt. or defined by trialist)
Non-alcoholic fatty liver disease, chronic viral hepatitis	1356.2 (n=6)	1425.5 (n=4)
All diseases/conditions	496.3 (n=764)	1052.0 (n=595)

Over time, the excess iron accumulates in these conditions, mainly as ferritin with a small amount as insoluble hemosiderin, most notably in the liver, eventually resulting in hepatic or cardiac complications unless treated. However, in view of the high variability in the mean serum ferritin concentrations noted in the cases with and without iron overload in Table 17.19, serum ferritin should not be used alone to detect iron overload. Instead, serum ferritin should be used alongside other iron biomarkers (e.g., transferrin saturation) and known genetic and clinical risk factors (Garcia-Casal et al., 2018).

Obesity may be associated with normal to increased concentrations of serum ferritin as a result of adiposity-related chronic inflammation. The cytokines, notably inter-leukin-6 (IL-6) and leptin, released from the enlarged adipocites, stimulate an increase in circulating hepcidin levels that block the release of iron from body stores into the circulation (independent of iron status), and down regulate intestinal iron absorption (Gartner et al., 2013). Overweight and obese women also have a higher Hb mass than normal weight women (Table 17.15), and thus a higher iron requirement (Cepeda-Lopez et al., 2019). Therefore, the impairment in iron absorption combined with an increased iron requirement may both contribute to the increased risk for iron deficiency in overweight and obese women, despite an apparently normal to increased concentration of serum ferritin (McClung and Karl, 2009; Cepeda-Lopez et al., 2019).

Certainly, in a study of Greek school children (9-13y; n=1493), lower values for serum iron and transferrin saturation and elevated serum sTfR concentrations indicative of iron-deficient erythropoiesis were observed in boys in the highest quartile of percentage body fat (determined via biolectrical impedance). In contrast, corresponding concentrations of ferritin were higher, compared to those in the lower quartile for body fat, as shown in Table 17.20. A similar trend was observed for girls (Moschonis et al., 2012). Clearly, the increased levels of serum ferritin in the obese children could have contributed to an underestimation of iron deficiency among this cohort, if the additional iron biomarkers had not been measured.

*Table 17.20. Biochemical and dietary indices of iron status across quartiles of percentage body fat mass in prepubertal children (mean values and standard deviations). TIBC, total Fe-binding capacity; TS, transferrin saturation. *Derived from ANOVA. †Mean values were significantly different between lower and higher quartiles after post-hoc multiple comparisons (P < 0·05, Bonferroni rule). ‡Mean values were significantly different between middle and higher quartiles after post-hoc multiple comparisons (P < 0·05, Bonferroni rule). §Mean values were significantly different between lower and middle quartiles after post-hoc multiple comparisons (P < 0·05, Bonferroni rule). From Moschonis et al. (2012).*

Percentage body fat mass	Lower quartile Boys (n=183)	Middle quartiles Boys (n=374)	Higher quartile Boys (n=183)	p*
Biochemical serum indices				
Fe (mg/L)	935† (339)	883‡ (341)	782†‡ (299)	0·001
TIBC (mg/L)	3388 (578)	3323 (497)	3421 (493)	0·144
TS (%)	27·8† (9·6)	26·8‡ (9·9)	23·2†‡ (9·1)	0·001
Ferritin (ng/ml)	27·3§† (16·0)	33·6§ (21·8)	36·3† (23·9)	0·001

Decreased erythropoiesis is associated with a decreased utilization of iron for Hb synthesis, so serum ferritin levels may be normal or slightly above normal. Decreased erythropoiesis may be associated with disease conditions such as chronic renal diseases and aplastic anemia.

Increased erythropoiesis leads to a decline in storage iron, reflected by a fall in serum ferritin levels owing to the increased demand for iron for Hb synthesis. Conditions in which increased erythropoiesis occurs include the stages of the life cycle when

iron needs are increased, such as during growth and pregnancy, after recovery from bone-marrow depression, hypoxia, and in response to vitamin B_{12} and folate therapy.

Genetic Hb disorders such as thalassemia traits and Hb EE may elevate serum ferritin (and sTfR) concentrations because of ineffective erythropoiesis and the short life span of red blood cells. As a consequence of these disturbances, levels of hepcidin are inappropriately low resulting in an excessive absorption of iron that leads to iron overload, and thus elevated serum ferritin levels(Manolova et al., 2019). As an example, in non-pregnant Cambodian women aged 18-45y (n=420), the geometric mean serum ferritin (and sTfR) concentration was significantly higher in women classified with Hb genotype EE compared to women with the normal Hb AA genotype (Karakochuk et al., 2015). Even during pregnancy, high serum ferritin concentrations have been reported among women with genetic Hb disorders such as sickle cell disease (Hb SS and Hb SC), some of whom had negligible or no iron in the bone marrow (Oluboyede, 1980). These findings highlight the difficulty of using serum ferritin (and sTfR) to assess the prevalence of iron deficiency in populations with a high prevalence of genetic Hb disorders.

Exercise when moderate, has little effect on serum ferritin concentrations, but if severe leads to an increase due to muscle damage and inflammatory reactions (WHO, 2007).

Malaria is associated with a high serum ferritin concentration irrespective of the parasite load. Several mechanisms are involved, including the destruction of red blood cells, the acute phase response, suppressed erythropoiesis, and the release of ferritin from damaged liver or spleen cells (WHO, 2007); see also “Malaria” in Section 17.2.1.

Acute and chronic liver disease may lead to abnormally high serum ferritin concentrations (Table 17.19), probably arising from the release of ferritin from damaged liver cells. Liver damage may also interfere with the clearance of ferritin from the circulation (Worwood, 1997). These elevated serum ferritin levels do not necessarily reflect a high intracellular concentration of ferritin, as indicated by the high mean serum ferritin in the absence of iron overload with liver disease in Table 17.19.

Leukemia and Hodgkin's disease also result in raised serum ferritin concentrations. In leukemia, these may be associated with: (a) increased deposition of iron in cells of the reticulo-endothelial system; (b) circulating leukemic cells containing high levels of ferritin; or (c) increased release of ferritin from damaged cells, as in liver disease. In Hodgkin's disease, the increased ferritin in the serum possibly comes from the lymphocytes (Worwood, 1979).

Other disease states such as alcoholism, liver cirrhosis, chronic hepatitis C, and rheumatoid arthritis are also associated with elevated serum ferritin concentrations (Table 17.19) (Garcia-Casal et al., 2018; Schutte et al., 2019). In some of these disease states, elevated serum ferritin values are observed even in cases of iron depletion as indicated by a bone marrow iron content of zero.

Pre-diabetes and Type 2 diabetes mellitus have been associated with elevated serum ferritin concentrations in several epidemiological studies (Cheung et al., 2013; Kunutsor et al., 2013; Yeap et al., 2015) although the underlying mechanism is poorly understood.

17.7.2 Interpretive criteria - use of the distribution of serum ferritin reference values

Reference value distributions for serum ferritin concentrations currently available have been drawn from a reference sample of unselected, apparently “healthy” individuals from the general population sampled during a nationally representative survey. Therefore, some of the persons in the sample will have almost no storage iron without being anemic, and a small proportion may even be anemic. As a result, even the “reference range” will include ferritin concentrations of persons who are iron deficient. This approach has been used in the following examples:

- U.S. Tables of mean, SEM, and selected percentiles (2.5^{th} to 97.5^{th}) for serum ferritin (µg/L) by age, sex, and race-ethnicity for all children and adults participating in U.S. NHANES III (2003-2006) (Pfeiffer et al., 2013); these are given in Appendix 17.4. Only breastfeeding infants and children were excluded. Modest, but significant decreases (≈10%) in serum ferritin concentrations were reported in women of child bearing age in U.S. NHANES III (2003-2006) compared to U.S. NHANES III (2001-2002)

- U.K. data compiled from the U.K. National Diet and Nutrition Surveys years 9-11 (2016-2017 and 2018-2019) includes methodological details. Tables of mean, median, and lower and upper 2.5 and 95th percentiles for serum ferritin (µg/L) by age and sex are included in a condensed ("zipped") comprehensive set of data tables based on information from (U.K. National Survey years 1-9)
- New Zealand Tables of mean (95th confidence interval) or mean (SEM) for serum ferritin values by age, sex, and race or ethnicity for children aged 5-14y in 2002 (Parnell et al., 2003) and persons 15y and older (2008-2009) in the New Zealand Adult Nutrition Survey (University of Otago and Ministry of Health, 2011).

17.7.3 Cutoffs for serum ferritin

Clearly, the assessment of the prevalence of iron deficiency based on a low serum ferritin concentration is challenging in settings where infection or inflammation are widespread. Many approaches have been proposed for adjusting for the effect of inflammation on ferritin and other iron biomarkers (Thurnham, 2014; Suchdev et al., 2017). Recently, a new regression modeling approach has been developed by the BRINDA group (Biomarkers Reflecting Inflammation and Nutritional Determinants of Anemia). In this approach, the inflammatory biomarkers (CRP and AGP) are treated as continuous variables because no clear cutoff was found for CRP or AGP at which there was a change in the relationship between inflammation and ferritin. By using this new approach, the full range and severity of inflammation can be accounted for, with greater corrections applied when the inflammatory biomarkers indicate severe inflammation; for more details see Raiten et al. (2015) and Suchdev et al. (2017).

The BRINDA approach differs from the earlier correction factor method developed by Thurnham (2014). In this earlier method, the stages of inflammation (incubation, early, late convalescence) were first classified, after which specific cutoffs for elevated levels of serum CRP (i.e.,> 5mg/L) and AGP (i.e.,> 1.0g/L) were applied; see Thurnham (2014) for more details. Table 17.21 compares the geometric means (95% CI) and the proportion with low serum ferritin (i.e., < 12µg/L) with and without adjustment for inflammation in a study of Indonesian infants age 12mos. Note the adjusted geometric mean value for serum ferritin was lower, and the proportion of infants at risk for iron deficiency (serum ferritin < 12µg/L) was higher, after applying the new BRINDA regression adjustment method (Diana et al., 2017). These findings are consistent with those reported elsewhere in preschool children and women of child-bearing age (Engle-Stone et al., 2017; Namaste et al., 2017). Furthermore, they emphasize the importance of measuring the two inflammatory biomarkers (CRP and AGP) recommended by WHO, and applying the BRINDA adjustment method prior to interpreting serum ferritin concentrations in settings where infection and inflammation co-exist with iron deficiency.

*Table 17.21. Impact of inflammation on serum ferritin concentrations of Indonesian infants at age 12mos. * Ferritin cutoff indicative < 12µg/L. From Diana et al. (2017).*

Biomarker in serum	Geometric mean (95% CI)	Proportion at risk (%)
Ferritin*: No adjustment	14.5µg/L (13.6-17.5)	44.9
Ferritin: Brinda adjustment	8.8µg/L (8.0-9.8)	64.9

Cutoffs for infants aged 4, 6, and 9 mos indicative of depleted iron stores and compiled by Domellöf et al. (2002a) are shown in Table 17.22.

Table 17.22. Suggested serum ferritin thresholds by age to classify individuals as iron deficient during epidemiological studies. [a]Data derived from studies of iron replete, breast-fed infants. [b]Data from NHANES III.

Age	Serum ferritin concentration (µg/L)	Reference
4mos	< 20	Domellöf et al. (2002a)
6mos	< 9	
9mos	< 5	
1-2y	< 10	Looker et al. (1997) [b]

Age	Serum ferritin concentration (µg/L)	Reference
3-5y	< 10	
6-11y	< 12	
12-15y	< 12	
≥ 16y	< 12	
0 - 23mos	< 12	WHO (2020)
24 - 59mos	< 12	
5 - < 10y	< 15	
10 - < 20y	< 15	
20 - 59y	< 15	
≥ 60y	< 15	
Pregnant women	< 15 first trimester	

They are based on serum ferritin concentrations set at −2SD for iron-replete Swedish breastfed infants. However, the rapid changes in iron status in the first year of life as fetal Hb is replaced by Hb A complicate the interpretation of serum ferritin during infancy.

Cutoffs for young children and adolescents ranging in age from 1y to > 16y differ slightly according to their source (Daru et al., 2017). Table 17.22 summarizes the cutoffs developed for five age groups by Looker et al. (1997) for use in the NHANES III surveys together with those recommended by WHO (2020) for infants, preschool children, school-aged children, and adolescents.

Concerns have been raised on the appropriateness of the cutoffs for preschool children shown in Table 17.22 in view of the limited data available to support these cutoffs (Daru et al., 2017). There is only one study (in apparently healthy Malawian children age 6-66mos) in which serum ferritin concentrations have been compared against the gold standard for iron deficiency (i.e., the absence of iron in bone marrow aspirates) (Jonker et al., 2014). Their results suggested that the current recommended cutoff for children (i.e., 12µg/L) although specific, identifying those persons who are genuinely well nourished, is not sensitive, with a poor ability to identify those persons with iron deficiency.

More recently, Mei et al. (2021) have used a population-based, physiological approach to define serum ferritin cutoffs in children age 12-59mos using two independent biomarkers of iron deficient erythropoiesis — Hb and sTfR. Their study was based on U.S. NHANES data over a 15y period. They suggest the use of a higher cutoff for children age 12-59mos (< 20µg/L) to determine the prevalence and distribution of iron deficiency in a population. A higher serum ferritin cutoff was also identified by Abdullah et al. (2017) in an earlier study of Canadian children (12-36mos) (i.e., 17.9µg/L) applying a similar physiological approach, although using only Hb and not sTfR.

Mei et al. (2021) emphasize that cutoffs defined using their physiological approach reflect the *onset* of iron-deficient erythropoiesis, whereas the values for preschool children and non-pregnant women shown in Table 17.22 are intended to characterize a more advanced stage of iron deficient erythropoiesis when bone marrow iron stores are exhausted. Consequently, the higher serum ferritin cutoffs of Mei et al. (2021), when applied, will reflect points where responses to iron depletion can be detected (Braat and Pasricha, 2021).

Cutoff for adults shown in Table 17.22 characterize an advanced stage of iron deficient erythropoiesis when bone marrow iron

stores are absent, based on limited evidence (Daru et al., 2017).

Mei et al. (2021) also applied their population-based, physiological approach to define ferritin cutoffs in non-pregnant women (15-49y) from the U.S. NHANES data. Again, they suggest a higher cutoff for serum ferritin for non-pregnant women (< 25µg/L) which reflects the *onset* of iron-deficient erythropoiesis, and which may be used to determine the prevalence and distribution of iron deficiency in a population.

Clearly, in view of the long-term adverse health outcomes that may be related to iron deficiency in children, and its impact on pregnancy outcome, more research to validate serum ferritin cutoffs using this new physiological approach in non-U.S. populations is urgently required.

Cutoffs for the elderly defined by WHO (2020) are shown in Table 17.22. There is some concern that they may need to be higher when used to diagnose iron depletion in this age group, although so far no consensus has been reached (Wawer et al., 2018).

Cutoffs for pregnancy are difficult to define in view of the numerous physiological changes that occur, including hemodilution and inflammation. WHO, based on a supporting Cochrane systematic review, failed to establish a cutoff for iron depletion during late pregnancy, because few pregnancy studies applied the gold standard bone marrow biopsy test (Garcia-Casal et al., 2018). As a consequence, WHO (2020) set a cutoff for serum ferritin to define iron deficiency only for the first trimester of pregnancy (i.e., < 15µg/L) as shown in Table 17.22.

Cutoffs for iron overload disorders based on elevated serum ferritin values are shown in Table 17.23.

Table 17.23. Cutoff values for serum ferritin (µg/L) indicative of iron overload in adults. From Pilch and Senti (1984).

Age group (y)	Males	Females
20-44	> 200	> 150
45-64	> 300	> 200
65-74	> 400	> 300

They were applied in the U.S. NHANES II and III in conjunction with a transferrin saturation > 70%.

WHO (2020), however, have specified cutoffs for serum ferritin for all apparently healthy individuals age ≥ 5y at risk of iron overload as > 200µg/L for males and > 150µg/L for females. They recommend using serum ferritin along with other biomarkers (e.g., transferrin saturation) and known genetic and clinical risk factors to diagnose iron overload. In this way, an incorrect diagnosis of iron overload when using serum ferritin alone will be avoided (McKinnon et al., 2013; Garcia-Casal et al., 2018).

WHO (2020) have also compiled cutoff values to define risk of iron overload in non-healthy individuals. For school-age children (5-12y), adolescents (13-19y), and adults of all ages, the suggested serum ferritin cutoff is > 500µg/L. However, WHO caution that a serum ferritin concentration > 500µg/L may indicate risk of iron overload or other diseases, and highlight the need for further clinical and laboratory evaluation to establish the diagnosis and underlying causes of the elevated ferritin levels.

Cutoffs for serum ferritin indicative of iron deficiency for individuals with infection or inflammation have been compiled by WHO (2020). For infants 0-23mos, preschool children (24-59mos), the WHO cutoff recommended is a serum ferritin < 30µg/L. For school-age children (5-12y), adolescents (13-19y), adults (20-59y), and older persons (≥ 60y), a serum ferritin < 70µg/L is recommended. However, many factors influence the effect of inflammation on serum ferritin, including the degree and duration of inflammation and immunity, so it is not unexpected that iron deficiency may exist in cases of inflammation even when serum ferritin concentrations are below these cutoffs (Garcia-Casal et al., 2018).

As an example, in the national survey in the Cameroon, applying a serum ferritin cutoff of < 30µg/L underestimated the prevalence of iron deficiency when compared to that based on inflammation-adjusted body iron stores < 0mg/kg (Engle-Stone et al., 2013). For more details on the assessment of body iron stores, see Section 17.10.1. Consequently, the preferred approach is to use serum CRP and AGP with the BRINDA method to adjust serum ferritin concentrations for the presence of inflammation.

WHO (2020) classify iron deficiency based on ferritin concentrations below their recommended cutoffs as severe, moderate, mild, or not a public health problem when the prevalence of iron deficiency is ≥ 40.0%, 20.0-39.9%, 5.0-19.9% or ≤ 4.9% respectively.

17.7.4 Measurement of serum ferritin

Venipuncture or capillary blood samples, fasted or nonfasted, can be used for serum or plasma ferritin, although within- and between-sample variation tends to be larger with capillary than venous specimens. Blood must be refrigerated immediately after collection, and processed within a few days. For ferritin assays, serum is stable for > 1wk at 4°C, and for > 1y at < –20°C, and can be subjected to 3 freeze-thawing cycles without any negative affect on the concentrations. For optimal long-term storage, serum must be stored at < –40°C, and when stored at –70°C, serum ferritin is stable for > 10y (Pfeiffer and Looker, 2017). Dried serum spot (DSS) samples can also be used (Ahluwalia et al., 2002), a major advantage for field studies. Before analysis, the DSS must be digested for 6h with cellulase from *Trichoderma reesei*.

In the past, serum ferritin was often determined with a single-incubation 2-site immunoradiometric assay (IRMA), a procedure used in the earlier U.S. NHANES and the U.K. Diet and Nutrition Surveys. In later U.S. NHANES (2004-2006), the Roche Tina-quant ferritin immunoturbimetric assay has been used on a clinical analyzer. This method, although automated, requires a large sample (> 150µL) which is a disadvantage for capillary or pediatric specimens. Today, commercial kits based on enzyme-linked immuno assays (ELISA) are also available which compare relatively well across manufacturers, with CVs ranging from 10—15% (Lynch et al., 2018). For a review of the performance and comparability of aboratory methods for measuring ferritin concentrations in serum or plasma, the reader is referred to Garcia-Casal et al., 2018. A ferritin certified reference material is available from WHO through the U.K. National Institute for Biological Standards and Controls (NIBSC, RM 94/572). This was used in NHANES III, and the use of this reference material as a calibration check is strongly encouraged.

This page titled 14.7: Serum ferritin (17.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.8: Erythrocyte protoporphyrin

When the supply of iron is adequate, heme is synthesized by the incorporation of iron into protoporphyrin IX by the enzyme ferro-chetalase. However, when there is insufficient iron to allow this reaction, zinc is substituted for iron, forming zinc protoporphyrin (ZPP). The ZPP then accumulates within the iron-deprived circulating erythrocytes, and can be measured directly in samples of whole blood using a portable hematofluorometer. An alternative chemical method can also be used whereby the porphyrins are extracted in a solvent, after which the porphyrins are measured fluorometrically. With this method, ZPP is converted to free erythrocyte protoporphyrin (FEP) (WHO, 2007). This extraction method now yields measurements almost identical to the direct measurements of ZPP performed using a portable hematofluorometer, provided the instrument used is properly calibrated and maintained, and a standardized procedure followed.

The zinc-for-iron substitution occurs predominantly within the bone marrow so a rise in the concentration of ZPP is one of the first indicators of insufficient iron in the bone marrow. Zinc protoporphyrin persists in the iron-deprived circulating erythrocytes for the duration of their lifespan (i.e., ≈ 120d). Therefore, an increased ZPP concentration in the blood indicates that the majority of the erythrocytes matured at a time when the iron supply was suboptimal (i.e., during the preceding 3-4mos), and thus provides a measure of uncomplicated iron-deficiency. The more pronounced the iron deficiency, the higher the ZPP values, with the highest values seen in severe iron deficiency anemia.

As ZPP and FEP are now interchangeable, the term erythrocyte protoporphyrin (EP) is now recommended, irrespective of the measurement method used, and is the term applied herein. A normal healthy person will have an EP concentration of less than 40-50µg EP/dL of red blood cells (RBCs), but when iron stores are exhausted (i.e., iron deficient erythropoiesis), the concentration may increase to more than 70-100µg/dL RBCs, reaching as high as 200µg/dL in iron deficiency anemia (WHO, 2007).

When a hematofluorometer is used to measure EP, Hb concentrations are also measured, allowing the EP to be expressed directly as a ratio of EP/Hb. Use of the ratio minimizes the effect of dilution that results from the expansion in plasma volume that occurs in pregnancy: EP, RBCs, Hb, and heme are diluted equally. In this way, a misinterpretation of laboratory results is avoided that may arise, for example, from plasma-volume changes in pregnancy. However, EP is expressed in several other ways in the literature, and may include EP relative to either whole blood (i.e., µg EP / dL whole blood) or RBCs (i.e., µg EP / dL RBCs). When the hematocrit is known, then the conversion of µg EP per dL whole blood to µg per dL RBCs is as follows:

$$\mu\text{g EP per dL red blood cells} = \frac{\mu\text{g EP per dL whole blood}}{\text{Hematocrit (\%)}}$$

Table 17.24. Comparison of the areas under the receiver operating characteristic (ROC) curves (Mean ± SE) of erythrocyte protoporp ciency in children (three to five years), non-pregnant women (15-49 years), and pregnant women from the National Health and Nutrit women, 1999-2006 for pregnant women), after excluding subjects who had signs of infection, as indicated by abnormal white blood possible liver disease, as defined by at least one of two abnormal elevations (more than two times the upper limit of normal value) o To generate the ROC curves, EP was converted to the preferred unit of µmol EP/mol heme by multiplying by 50. Values with differen under the curves. From Mei et al. (2017).

	Children (n=762) [2]	Non-Pregnant Women (n=2746) [2]
EP	0.785 ± 0.049[a]	0.864 ± 0.011[a]
Hb	0.596 ± 0.054[b]	0.841 ± 0.011[a]
MCV	0.703 ± 0.052[a]	0.802 ± 0.014[b]

From µg EP per dL RBCs to µg EP per g Hb: multiply by 0.037. From µg EP per dL RBCs to µmol EP per mol heme: multiply by 0.87. These factors are based on the assumption that the MCHC is normal, although when MCHC is measured in individual samples, then an appropriate factor can be calculated. For more details on the differing EP reporting units and their inter-conversion, see Beard in WHO (2007).

Table 17.24 compares the receiver operating characteristic (ROC) curves used to characterize the sensitivity and specificity of

EP, Hb, and MCV measurements in screening iron deficiency based on data from U.S. NHANES III (Mei et al., 2017). For children aged 3-5y, EP expressed as µmol/L (measured by the modified acid extraction method) was a better screening tool for iron deficiency than Hb, but not statistically different from MCV. For non-pregnant and pregnant women, the performance of EP and Hb were similar, with both being superior to MCV. Iron deficiency was defined as total body iron stores (TBI) < 0mg/kg body weight; for details on TBI, see "Body iron"" in Section 17.10.1.

Nevertheless, the results presented in Table 17.24 apply to U.S. population groups for whom iron deficiency is an important cause of anemia and may not necessarily apply to populations with a high prevalence of infectious diseases or for persons with other non-nutritional causes of iron deficient erythropoiesis.

Table 17.25. Measured ratios of EP:Hb, their interpretation, and recommendations.
EP, erythrocyte protoporphyrin; Hb, hemoglobin; CBC, complete blood count; ACD, anemia of chronic disease; APP, acute phase prot

Measurements	Interpretation	Recommenda
Low EP/Hb ratio (< 60µmol/mol)	Adequate systemic iron supply	Iron stores car overload (TSAT
Mid-range EP/Hb ratio (60-80µmol/mol)	Possible non-replete iron status; consider inadequate diet, ACD, or other causes	CBC may supp blockade; coul
High EP/Hb ratio, (> 80µmol/mol) and low ferritin concentration (< 20µg/L)	Iron deficient erythropoiesis attributable to low marrow iron supply, may be related to depleted iron stores	Iron suppleme
High EP/Hb ratio (> 80µmol/mol) and high ferritin concentration (> 200µg/L)	Severe inflammatory blockade, ACD, other causes of impaired iron utilization	Correct the cau

Table 17.25 provides guidelines on the interpretation of EP/Hb ratios in the diagnosis of nutritional and non-nutritional causes of iron deficient erythropoiesis compiled by Labbé and Dewaanji (2004) and reproduced by Beard (WHO, 2007).

17.8.1 Factors affecting erythrocyte protoporphyrin

Biological variation for EP is ≈6.5%. This includes circadian variation (day-to-day) and diurnal variation (within one day). Levels between 18.00hrs and midnight are higher for reasons that are unkown (Soldin et al., 2003).

Age-related trends suggest that levels of EP are highest from age 0-12mos, probably reflecting the rapid erythropoiesis in this age group. After infancy, they decrease until adolescence (10-17y), when they appear to increase, declining slightly thereafter (Soldin et al., 2003; Table 17.26).

Sex-related changes in EP have been described, with girls having significantly higher levels than males at 10-17y. Likewise, adult women have slightly higher levels than males (Soldin et al., 2003).

Iron deficient erythropoiesis, characterized by the exhaustion of iron stores, leads to a slow but progressive increase in EP in the peripheral blood which is apparent before any decline in Hb and MCV. The increase in EP concentration occurs slowly because only the newly produced erythrocytes containing ZPP appear in the peripheral blood to replace those erythrocytes at the end of their life span (i.e., ≈ 120 days) (Mei et al., 2017).

Infection / Inflammation is associated with elevated levels of EP. This is due to the increase in pro-inflammatory cytokines in the systemic circulation which stimulate an increase in hepcidin concentrations. Elevated level of circulating hepcidin reduce both iron absorption and the release of iron from body stores so that iron is no longer available for transport to the bone marrow for erythropoiesis. When the iron is no longer available, divalent zinc is inserted into protoporphyrin, so concentrations of EP are elevated, irrespective of the EP assay method used. As a result, during inflammation, assay of EP alone cannot be used to distinguish between the anemia of iron deficiency and the anemia of chronic disease.

Pregnancy is believed not to result in a decline in EP when expressed as a molar ratio to Hb, as the ratio is theoretically independent of hemodilution. Consequently, EP expressed as a molar ratio to Hb has been proposed as a preferred marker of iron status in pregnancy (Schifman et al., 1987).

However, in pregnant women in Kenya without evidence of inflammation or infection with *Plasmodium* or HIV, EP reportedly had a limited ability to discriminate between women with and without iron deficiency even when combined with Hb, and irrespective of whether EP was measured in washed erythrocytes or whole blood using a hematofluorometer or by solvent extraction. Moreover, the prevalence of iron deficiency was grossly over-estimated when applying the conventional EP cutoff for whole blood (> 70µmol/mol Hb) in comparison with that based on plasma ferritin concentrations < 15µg/L, especially when the prevalence of iron deficiency was low (e.g., 10%) (Mwangi et al., 2014). Based on these results, the usefulness of EP:Hb molar ratio to assess iron status during pregnancy, particularly in disadvantaged settings, remains uncertain.

Lead toxicity can result in elevated concentrations of EP due to disturbances in the heme synthesis pathway; lead inhibits porphobilinogen synthase which results in an increased excretion of aminolaevulinic acid (ALA) in the urine. Lead may also have a direct effect on the enzyme ferrochetalase (WHO, 2007). As a consequence, ZPP (now termed EP) is formed instead. In U.S. NHANES II, EP values were elevated in children 1-4y with high blood lead values, although lower in U.S. NHANES III following the introduction of lead-free petrol (Looker et al., 1997).

Genetic Hb disorders such as sickle cell anemia, thalassemia, and Hb E increase EP concentrations, most likely due to ineffective erythropoiesis and red blood cells with a short life span (Graham et al., 1996; Mwangi et al., 2014; Pfeiffer and Looker, 2017). The relative iron deficiency that occurs in these disorders has also been proposed as a possible factor, although seemingly unlikely because of the increase in EP reported in individuals with these conditions, even in the absence of nutritional iron deficiency (Graham et al., 1996).

Malaria from *Plasmodium* infection is associated with elevated EP concentrations in whole blood or erythrocytes, as a consequence of impaired erythropoiesis and the induction of the acute phase response, and subsequent upregulation of hepcidin; see also “Malaria” in Section 17.2.1 for more details.

17.8.2 Interpretive criteria — use of the distribution of erythrocyte protoporphyrin reference values

Details of the distribution of reference values from four sources are described below, of which only one was based on a “healthy” reference sample.

- U.S. Tables of mean, SEM, selected percentiles for erythrocyte protoporphyrin (measured as FEP) expressed as µmol/L RBCs by age, sex, and race based on all persons in U.S. NHANES II.
- New Zealand Tables of mean, SEM for the ZnPP: Hb ratio (µmol/mol) by age, sex, and race or ethnicity for all persons > 15y participating in the 1997 national survey; those with conditions affecting ZPP levels were not excluded. Measurements of ZPP were performed using a hematofluorometer Russell et al., 1999).

Table 17.26. Pediatric reference ranges of ZPP blood concentrations (2.5th-97.5thpercentiles µg/dL) for females and males age 0-17 (2003).

	Blood ZPP Reference Ranges 2.5th-97.5th percentiles (µg/dL)*		Blood ZPP Ranges 2.5
Age Groups	**Males**	**Females**	**Males**
0-12mos n=145M; 203F	8.5-34.5	9-40	15.6-63.5
13-24mos n=605M; 725F	10-34	11-32	18.4-62.6

Age Groups	Blood ZPP Reference Ranges 2.5th-97.5th percentiles (µg/dL)* Males	Females	Blood ZPP Ranges 2.5 Males
2-5y n=1,926M; 1,822F	5-35	10-31	9.2-64.4
6-9y n=522M; 408F	6-31	9-30	11.0-57.0
10-17y n=61M; 61F	5.5-31.5	5-33.5	10.1-58.0

- Canadian Tables of geometric mean (95th CI) and reference ranges (2.5th-97.5th percentile) for ZPP (expressed as µg/dL whole blood and as µmol/mol heme) by sex and age for infants and children age 0-17y; children known to have conditions affecting ZPP levels were not excluded. Data were from blood samples taken in a large pediatric hospital (n=6478) and assayed using a hematofluorometer (Soldin et al., 2003). Table 17.26 presents the pediatric "reference range" or "range of normal" for males and females for five age groups, with the 97.5th percentile value corresponding to the upper reference limit. Values above the upper reference limit are classified as "unusually high".
- U.S. Tables of mean, SEM, selected percentiles for erythrocyte protoporphyrin (as FEP expressed as µmol/L) by age, sex, and race for a "healthy" reference sample. Data from persons with conditions known to affect iron status were excluded. Measurements were performed using a variant of the solvent extraction method (Looker et al., 1997).

17.8.3 Cutoffs for erythrocyte protoporphyrin

Cutoff for infants aged 4, 6, and 9mos expressed as µmol/mol heme based on iron-replete breastfed infants are presented in Table 17.16 (Domellöf et al. 2002). WHO (2017) present a single cutoff for infants 0-12mos as > 80µg/dL RBC (i.e., > 1.42µmol/L RBC; 2.7µg/g Hb; > 70 µmol/mol heme) (Table 17.27).

Cutoffs for children ages 1-2y, 2-9y and adolescents (girls) age 10-17y are also shown in Table 17.27. These cutoffs are based on the CDC recommended demarcation between iron deficient erythropoiesis and iron sufficient erythropoiesis, expressed as µmol/L RBCs or as µg/dL RBCs. Table 17.27 also includes the 97.5th percentiles (expressed as µg/dL RBCs) adapted from Soldin et al. (2003) for comparison. Note the large discrepancy between the CDC cutoffs and the statistically derived upper reference limit based on the 97.5th percentile of Soldin et al. (2003), emphasizing that the derivation of cutoff values is very method dependent.

Table 17.27. The threshold concentrations of erythrocyte protoporphyrin by age groups at which iron-deficient erythropoiesis occurs the Centers for Disease Control and Prevention (CDC) recommendations and the 97.5th percentile of the distribution of values in he
[a]CDC recommendation as the demarcation between iron-deficient erythropoiesis and iron-sufficient erythropoiesis. Abstracted from

Age	EP (µmol/L RBC	Suggested threshold[a] (µg/dL red cells)	97.5th percentile (µg/dL red cells)
0-12mos	> 1.42	> 80	40
1-2y	> 1.24	> 80	32
2-9y	> 1.24	> 70	30

Age	EP (µmol/L RBC	Suggested threshold[a] (µg/dL red cells)	97.5th percentile (µg/dL red cells)
10-17y (girls)	> 1.24	> 70	34

WHO have provided cutoffs for the detection of iron deficiency for children < 5y and 5y or more as > 61µmol/mol heme (i.e., > 70µg/dL RBC) and > 70µmol/mol heme, (i.e., > 80µg/dL RBC) respectively. These are based on the sensitivity and specificity for detecting the absence of storage iron (WHO, 2001).

Cutoff for adults for EP in whole blood of > 70µmol/mol heme (i.e., > 80µg/dL RBC; > 2.7µg/g Hb) was developed by Looker et al. (1997). This cutoff approximated the 95th percentile of the U.S. NHANES III "healthy" reference sample. WHO (2001) suggest that this cutoff for adults is expressed in one of the following three ways: > 70µmol/mol heme; > 80µg/dL RBC; >≈ 3.0µg/g Hb.

Cutoff for pregnancy is the same as that used for EP for non-pregnant adults i.e., in whole blood: > 70µmol/mol heme (> 80µg/dL RBC, >≈ 3.0µg/g Hb). and compiled by Looker et al. (1997). The cutoff for non-pregnant adult women was selected because EP, when expressed as the ratio of EP/Hb, was believed to remain stable during pregnancy, as noted earlier. However, the use of this conventional cutoff for pregnancy has been questionned. Mwangi et al. (2014) reported that applying this cutoff (> 70µmol/mol heme) led to a gross overestimate of the prevalence of iron deficiency in pregnancy in their study of Kenyan women.

17.8.4 Measurement of erythrocyte protoporphyrin

Two methods can be used to measure erythrocyte protoporphyrin: a direct and indirect method. The direct method uses a portable hematofluorometer to measure the fraction of blood protoporphyrin that exists as the zinc chelate (i.e., ZPP). The manufacturer of the hematofluorometer assumes that ZPP comprises 95% of all the erythrocyte protoporphyrin in blood, thus providing the basis for the estimates of EP concentrations (WHO, 2017). Only a drop of fresh whole, unprocessed blood (< 20µL) obtained by capillary sampling is required, and the measurement is available in ≈ 1min. Calibration standards are normally provided to check the stability of the hematofluorometer (Labbé et al., 1999), although for recalibration, the equipment must be shipped back to the manufacturer. A NCCLS Guideline for erythrocyte protoporphyrin testing is available.

Several issues may affect the direct measurement of EP using a hematofluorometer. These include the requirement to oxygenate the blood sample if the time between the blood sampling and EP measurement is > 30min, and the need to wash the red blood cells with saline. Incomplete oxygenation produces falsely low values because of a shift in Hb absorption. To ensure complete oxygenation and dissolution of red-cell aggregates, the whole blood must be stirred gently using a disposable plastic pipet tip. Wooden applicator sticks should not be used for mixing as they may contain fluorescent by-products. Successive observations should be within 10%. Use of red blood cells that are washed with saline rather than whole blood are recommended to remove interference by non-porphyrin fluorescent compounds such as bilirubin, certain drugs, and high riboflavin concentrations that may be present in the plasma (Hastka et al., 1992), but is rarely performed. Application of these strategies is said to reduce the number of falsely high results and improve the reproducibility and precision of the EP measurements; they are discussed in detail in Labbé et al. (1999).

The indirect method involves a variation of the acid extraction method of Chisolm and Brown (1975) during which ZPP is converted to FEP. This method is more time consuming and less widely used, although used by U.S. NHANES II and III (Mei et al., 2017). It involves the extraction of total protoporphyrin from anticoagulated whole blood into a mixture of ethyl acetate-acetic acid, then back-extraction into dilute hydrochloric acid, followed by measurement of protoporphyrin fluorometrically. Measurements of EP using this method can be made on stored blood samples or from blood collected on filter paper in the field. Results for EP using this indirect method are similar to those using the direct method, as noted earlier, provided the hematofluorometer used is properly calibrated and maintained and a standardized procedure is followed (Mei et al., 2003).

A novel non-invasive method has been developed to optimally measure EP in the microcirculation of the lower lip by using an optical fiber probe to illuminate the lip and acquire fluorescence emission spectra in ≈ 1min. The method is fast and provides specific testing for iron deficiency and iron deficiency anemia at the point of care without the need for blood drawing; details are given in Hennig et al. (2016). Method reliability was compared by measurements of EP by both HPLC and fluorescence spectroscopy, and feasibility tested on young children. Measurements of EP by this novel method to identify iron deficiency and iron deficiency anemia in children 9mos-5y compared well with those based on ferritin, serum soluble transferrin receptor, and

Hb (Homann et al., 2019).

This page titled 14.8: Erythrocyte protoporphyrin is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.9: Serum soluble transferrin receptor (17.9)

Transferrin receptor (TfR) is an iron-related protein that regulates the uptake of transferrin, carrying iron into body cells. The surfaces of all cells express TfR in proportion to their requirement for iron, although the largest number are in the erythron, placenta, and liver (Beguin, 2003). A soluble form of sTfR circulates in human plasma bound to transferrin and is usually called soluble transferrin receptor (sTfR). The predominant source of sTfR in the plasma pool are the cells of the developing red-cell mass — the erythroblasts and reticulocytes. Hence, the concentration of sTfR reflects erythropoietic activity in the bone marrow (Table 17.28).

Table 17.28. Serum transferrin receptor concentration changes in human disease. Based on Worwood (2002).

sTfR	Condition
Increased or not affected	Increased erythroid proliferation or ineffective erythropoiesis:
	— Autoimmune hemolytic anemia
	— Hereditary spherocytosis
	— Beta thalassemia/HbE
	— Hb H disease
	— Sickle cell anemia
	— Polycythemia vera
	—Vitamin B_{12} deficiency
	— Folic acid deficiency
	Decreased tissue iron stores:
	— Iron-deficiency anemia
Normal to increased	Idiopathic myelofibrosis
	Myelodysplastic syndrome
	Chronic lymphocytic leukemia
Normal	Hemochromatosis (see text)
	Acute and chronic myeloid leukemia
	Most lymphoid malignancies
	Solid tumors
Decreased	Chronic renal failure
	Aplastic anemia
	Post bone-marrow transplantation
	Hypertransfusion

In the first stage of the development of uncomplicated iron deficiency anemia, termed “iron depletion” (Section 17.1.9), the supply of iron to the functional compartment is not compromised, so concentrations of sTfR in serum or plasma are unaffected. However, in the second stage, termed “iron-deficient erythropoiesis” when body iron stores are exhausted and the availability of iron to tissues is compromised, transferrin production is upregulated to increase iron transport, and cell surface TfR1 expression increases. This allows the cells to compete more effectively for transferrin-bound iron, and results in an increase in sTfR concentrations in plasma or serum. In contrast, when there is sufficient body iron, TfR1 is down regulated so concentrations of sTfR in the plasma or serum fall. As a result, sTfR levels in plasma or serum reflect the functional iron deficit (Gimferrer et al., 1997). For example, in elderly women with uncomplicated iron deficiency, but not necessarily anemia, sTfR levels in plasma were 1.3 times the levels in iron-sufficient controls (Ahluwalia et al., 1995), whereas in those with microcytic, hypochromic anemia, sTfR levels

in serum were 5.8 times that of the controls (Gimferrer et al., 1997). Concentrations of sTfR in serum have also been shown to correlate with iron status based on stainable bone marrow iron in several studies (Chang et al., 2007; Koulaouzidis et al., 2009).

One of the main advantages of using serum sTfR is that concentrations are less strongly affected by inflammation than serum ferritin. Nevertheless, because any change in the rate of erythropoiesis will in turn affect serum sTfR concentrations, sTfR can only be used as an indicator of iron status when iron stores are exhausted and there are no other causes of abnormal erythropoiesis (WHO, 2007). Table 17.28 depicts some of the disease conditions that are known to alter serum sTfR concentrations. More details of factors affecting serum sTfR concentrations are outlined below.

17.9.1 Factors affecting serum soluble transferrin receptor

Biological variation for soluble transferrin receptor concentration is low, with reasonably stable concentrations within an individual so that only one blood sample is sufficient to measure sTfR. Cooper and Zlotkin (1996)determined the total day-to-day within-subject coefficient of variation (CV) for a group of healthy men (n=10) and women (n=11) aged 19-46y, based on measurements on 10 nonconsecutive days over a 4wk period. Biological variation for serum sTfR in this study was 12.5%, compared with 8.9% (Lammi-Keefe et al., 1993) and 8.1% (Belza et al., 2005) in other population groups.

Age-related changes in serum sTfR occur, in part as a result of the increase in erythropoiesis during growth, although other mechanisms may be involved. Levels increase from birth until ≈ 7.5mos in breast-fed infants, after which they level off, remaining relatively stable for the rest of the first year of life (Ziegler et al., 2014).

During childhood and adolescence, data to define age related trends in serum sTfR levels are more limited. In the U.S. NHANES 2003-2010 surveys, serum sTfR concentrations were higher in preschool children age 1-2y than those aged 3-5y (Figure 17.14).

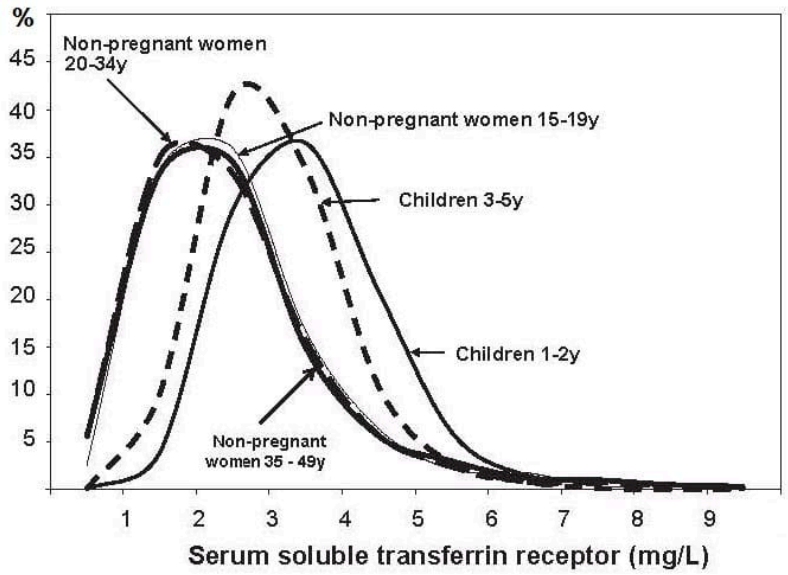


Figure 17.14. Distributions of serum soluble transferrin receptor (sTfR) concentrations in U.S. preschool children and non-pregnant women of childbearing age participating in the U.S. NHANES 2003-2010. Redrawn from Mei et al. (2012).

In adolescent girls (15-19y), concentrations were higher than those for adult women aged 20-34y, although not higher compared to adults aged 35-49y (Mei et al., 2012). These age-related trends have also been reported in other studies (Choi et al., 1999; Virtanen et al., 1999).

Sex-related differences in serum sTfR are small (Worwood, 2002). During early infancy, values may be higher for males than for females, perhaps as a result of sex differences in fetal iron accretion (Domellöf et al., 2002a). These differences have not been

reported consistently in older infants (9-15mo) (Yeung and Zlotkin, 1997). In the U.S. NHANES 2003-2010 survey, however, sTfR concentrations were also higher in preschool-aged boys than girls after adjusting for socio-demographic variables (Mei et al., 2012).

Race may influence serum sTfR concentrations. Concentrations for African American pre-school children and non-pregnant women were 9% higher than those for Caucasians, Asians, and Hispanics in the U.S. NHANES 2003-2010 surveys (Mei et al., 2012). Such race-related differences may be attributed to the well-established but as yet unexplained lower Hb concentration known to occur among persons of African descent(Allen et al., 1998).

Tissue iron deficiency leads to increases in serum sTfR, as noted earlier. Even after iron stores are totally exhausted, serum sTfR concentrations continue to increase in proportion to the increasing tissue iron deficit, as Hb concentrations fall (Skikne et al., 1990). This trend has been observed in neonates (Rusia et al., 1996), infants and young children (Olivares et al., 2000), pregnant women (Carriaga et al., 1991), and adults (Skikne et al., 1990). When the iron supply is not limiting, then concentrations of sTfR in plasma or serum can be used to monitor bone marrow erythropoiesis (Lynch et al., 2018).

Iron overload in patients with hemachromatosis and in those with African iron overload has been associated with serum sTfR concentrations within the normal range (Baynes et al., 1994), so that measurement of sTfR is not useful for diagnosing iron overload.

Pregnancy is accompanied by marked physiological changes that may confound the interpretation of serum sTfR levels. Although reported to provide a sensitive biomarker of iron deficiency during pregnancy (Carriaga et al., 1991), sTfR levels are also influenced by changes in the rate of erythropoiesis during this time. For example, during early pregnancy, the rate of erythropoiesis decreases and may mask the development of iron deficiency (Akesson et al., 1998), whereas in later pregnancy when increases in serum sTFR are observed, whether they are related to the increase in erythropoiesis rather than iron depletion is uncertain (Worwood et al., 2017). The elevated concentrations of serum sTfR that are evident in later pregnancy return to non-pregnant values 12wks after delivery (Choi et al., 2000).

In pregnant women in Malawi in which iron biomarkers were investigated by comparison with stained bone marrow aspirates, sTfR did not enhance the sensitivity and specificity for detecting iron deficiency anemia in pregnancy (Van Den Broek et al., 1998). Lynch (2011) has proposed the use of the ratio of sTfR to serum ferritin to detect iron deficiency during trimesters 1 and 2, but has no recommendation for the third trimester of pregnancy. However, total body iron (Section 17.10.1) was used to assess iron status in women in the 1st, 2nd, and 3rd trimester of pregnancy in the cross-sectional U.S NHANES 1999-2006 study (Mei et al., 2011).

Table 17.29 presents the changes in several iron biomarkers, including serum sTfR, body iron (Section 17.10.1) and serum hepcidin (Section 17.11.1), in a longitudinal study of pregnant women in the Gambia at 14wk, 20wk, and 30wk gestation, supplemented with iron from enrolment to delivery (Bah et al., 2017).

Table 17.29. Iron and hematologic indexes in pregnant Gambian women. Values are the arithmetic (hemoglobin, MCV, sTfR, CRP, and total body iron) or geometric (ferritin, sTfR-F index, and hepcidin) means (ranges). P values are for 2-sided paired t-tests comparing analytes between 14 and 20wk and between 20 and 30wk of gestation. CRP, C-reactive protein; MCV, mean cell volume; sTfR, soluble transferrin receptor; sTfR-F index, sTfR/log$_{10}$ (ferritin). From Bah et al. (2017).

Index	14wk n=395	20wk n=375	P	30wk n=367	P
Gestational age, wk	14.1 (8.0, 21.3)	20.4 (15.0, 26.9)		30.5 (25.4, 34.4)	
Hemoglobin g/dL	11.55 (7.2, 17.9)	11.00 (7.4, 14.5)	< 0.001	10.77 (6.2, 14.4)	< 0.0001
MCV fL	81.9 (60.9, 98.6)	83.8 (62.9, 104.0)	< 0.001	83.6 (61.8, 99)	NS
Serum ferritin µg/L	20.69 (0.1, 237.2)	19.20 (0.1, 273.8)	NS	14.29 (0.1, 315.5)	< 0.001

Index	14wk n=395	20wk n=375	P	30wk n=367	P
Serum sTfR mg/L	4.41 (0.58, 17.97)	4.25 (1.12, 15.42)	NS	4.80 (1.49, 17.81)	< 0.001
Serum sTfR-F index	3.29 (−14.80, 111.63)	3.26 (−5.63, 56.43)	NS	3.86 (−87.19, 76.26)	NS
Serum CRP mg/L	1.70 (0.02, 43.48)	2.41 (0.00, 59.32)	0.001	2.29 (0.01, 126.68)	NS
Total body iron mg/kg	2.70 (−18.38, 14.40)	2.50 (−17.57, 14.06)	NS	1.18 (−17.84, 11.02)	< 0.001
Serum hepcidin ng/L	1.59 (0.03, 49.79)	1.23 (0.02, 45.70)	0.006	1.09 (0.04, 135.69)	NS

Changes in these iron biomarkers were marked. Note that the Hb declined from 14wk through 30wk gestation, hepcidin declined from enrolment through 20wk gestation, whereas the changes for serum ferritin, serum transferrin receptor, and thus total body iron were only significant here between 20wk and 30wk gestation. Extensive investigations of associations between the variables revealed that hepcidin was superior to sTfR (and Hb) as an indicator of iron deficiency; see Section 17.11.1 for more discussion of hepcidin.

Infection / Inflammation is now known to increase serum sTfR concentrations, although less strongly than for serum ferritin (Rohner et al., 2017). In preschool children and women of reproductive age, the prevalence of elevated sTfR concentrations (i.e., > 8.3mg/L) indicative of iron-deficient erythropoiesis decreased incrementally as serum CRP and AGP deciles decreased, based on a large dataset from the BRINDA Project (Engle-Stone et al., 2017). The decrease observed was more pronounced for AGP, a measure of more longer-term exposure to inflammation than CRP, as shown in Figure 17.15.

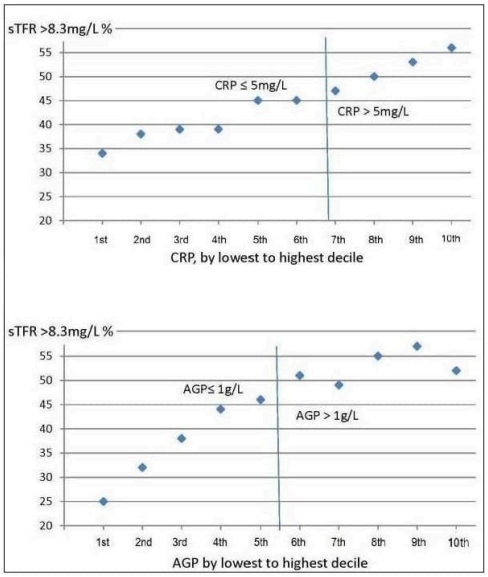


Figure 17.15. Prevalence of elevated sTfR concentrations in preschool children by CRP and AGP deciles: BRINDA project. Analysis was restricted to surveys (Bangladesh, Cameroon, Côte d'Ivoire, Kenya 2007, Kenya 2010, Laos, Liberia, Philippines, and Papua New Guinea) that measured both CRP and AGP for comparability between CRP and AGP relations with biomarkers (n=9326). AGP, α-1-acid-glycoprotein; BRINDA, biomarkers Reflecting Inflammation and Nutrition Determinants of Anemia; CRP, C-reactive protein; sTfR, soluble transferrin receptor.

Redrawn from Rohner et al. (2017).

As a result, measurement of AGP has been recommended in any procedure used to adjust serum sTfR for the presence of inflam-

mation. Details of the BRINDA regression approach to adjust serum sTfR values for inflammation are available in Rohner et al.(2017). However, this approach needs to be evaluated in different settings and in pregnant women (Gupta et al., 2017).

Other disease conditions also affect sTfR concentrations (see Table 17.28), which must be considered when using sTfR as an indicator for nutritional iron deficiency(Rohner et al., 2017).

Malaria has the potential for influencing serum sTfR concentrations in two opposing ways. A decrease in serum sTfR concentrations may result from a reduction in erythropoiesis associated with the marked inflammatory response during the acute phase of malarial infections. Erythropoiesis may also be suppressed in part by a direct inhibitory effect of hemozoin, a parasite by-product of Hb (Lamikanra et al., 2007).

Alternatively, serum sTfR concentrations can be elevated in malaria after the initial acute infection. This trend is thought to be due to an increase in erythropoietic activity to compensate for the increased red-cell destruction. Hence, in areas where both iron deficiency and malaria occur, the reliability of serum TfR as an indicator of iron status is uncertain. Rohner et al. (2017) reported a decrease in the prevalence of iron-deficient erythropoiesis based on elevated serum sTfR concentrations, after applying the regression approach based on AGP to adjust for the presence of malaria. Nevertheless, these investigators stressed that additional adjustments for malaria, unrelated to the acute-phase response, may also be needed.

Genetic Hb disorders such as thalassemia traits and Hb EE may elevate serum sTfR due to ineffective erythropoiesis (i.e., an increase in the proportion of immature red cells destroyed within the bone marrow) and red blood cells with a short half-life (Khatami et al., 2013; Manolova et al., 2019). Women and children in Cambodia, for example, identified with Hb EE had serum sTfR concentrations that were significantly higher compared to those with either a normal Hb or Hb AE (Karakochuk et al., 2015; George et al., 2013). Clearly, it is difficult to use serum sTfR to detect tissue iron deficiency in populations where the prevalence of these disorders is high.

Increased or ineffective erythropoiesis in disease conditions itemized in Table 17.28, are associated with elevated serum sTfR concentrations, as reported earlier. Note that both folic acid and vitamin B_{12} deficiency are included because they are both associated with defective but elevated erythropoiesis. In such cases, use of serum sTfR measurements alone would yield a false-positive diagnosis for iron deficiency.

Decreased erythropoiesis associated with several disease conditions listed in Table 17.28 leads to a fall in serum sTfR concentrations (Beard et al., 1996; Ahluwalia, 1998). Fortunately, these conditions can usually be identified by measuring serum ferritin, which is normal in hematologic disorders with decreased erythropoiesis (Section 17.7.1).

High altitudes may result in higher serum sTfR concentrations, as a consequence of the increase in the total body erythroid mass in response to systemic hypoxia. Hypoxia increases the expression of erythropoietin, and the erythropoietin drives the production of new erythrocytes (Tanno and Miller, 2010).

Lifestyle factors such as smoking have been linked to a significant decrease in serum sTfR concentrations (Raya et al., 2001; Pynaert et al., 2008). In contrast, sTfR levels appear to be unaffected by alcohol consumption, despite the liver damage and change in iron metabolism that occurs (Speeckaert et al., 2010). High levels of physical activity are associated with elevated levels of sTfR. The etiology is uncertain but possibly linked with a lower iron status (Woolf et al., 2009).

Obesity has been linked to elevated serum sTfR concentrations. This trend has been related to elevated levels of hepcidin in response to obesity-related inflammation and an increase in Hb mass (i.e., total Hb in g) in overweight and obese women (see Table 17.15 and Section 17.6.1). Together these two factors contribute to an increased risk of iron deficiency in obesity Cepeda-Lopez et al., 2019).

Oral contraceptive agents and hormone replacement therapy do not affect serum TfR concentrations (Raya et al., 2001).

17.9.2 Interpretive criteria — use of the distribution of soluble transferin receptor reference values

Defining reference values for sTfR in healthy individuals is difficult because serum sTfR concentrations and the units of expression vary with the assay; even different kits based on the ELISA technique are associated with different “reference” ranges. An international standard is now available (Thorpe et al., 2010), so distributions of reference values and internationally applicable cutoff values for sTfR should be available in the future.

Table 17.30. Geometric mean, 2.5th and 97.5th percentile (95% confidence interval) for serum soluble transferrin receptor (sTfR, mg/L) in the U.S. population of U.S. children aged 1-5y and non-pregnant women aged 15-49y participating in the U.S. NHANES III 2003-2010. The following exclusions were applied: children with anemia, iron deficiency, infection, inflammation and liver disease. † Within a group, values with different superscripts are significantly different in means, $p < 0.05$ (2-tailed t-test) and a Bonferroni adjustment was used for to correct the p values for the number of comparisons across each demographic characteristic in sTfR. Data from Mei et al.(2012).

	n	Geometric mean†	2.5th percentile	97
Children				
1-2y M	386	4.29 (4.16, 4.42)[a]	2.53 (2.22, 2.94)	6.4
1-2y F	354	4.19 (4.07, 4.31)[a]	2.76 (2.66, 2.96)	6.0
3-5y M	735	3.91 (3.84, 3.99)[b]	2.55 (2.40, 2.64)	5.9
3-5y F	678	3.83 (3.76, 3.90)[b]	2.57 (2.38, 2.71)	5.6
Non-pregnant women				
15-19y	1237	3.14 (3.06, 3.23)[a]	1.86 (1.77, 1.97)	5.2
20-49y	3423	3.05 (3.00, 3.09)[b]	1.65 (1.59, 1.72)	5.4

Details of the distribution of reference values from three sources are described below, each drawn from a "healthy" reference sample in which participants with conditions known to affect iron status were excluded. For each distribution of reference values, a different assay method was used.

- U.S. Tables of geometric mean, median (not shown), and lower (2.5th) and upper (97.5th) percentiles for "healthy" preschool children aged 1-5y (by age and sex) and non-pregnant women 15-49y (by age) participating in the U.S. NHANES III (2003-2010) (Table 17.30). Geometric means were calculated because serum sTfR values were positively skewed. Measurements of serum sTfR were performed using the Roche Tina-quant automated sTfR immunoturbidity assay. Correlation between this immunoturbidity assay and the ELISA (Ramco Laboratories) assay is reportedly good, although values from the Roche assay are on average 30% lower than those with the ELISA assay (Pfeiffer et al., 2007).
- Korean table of mean (SD), 95% CI for "healthy" subjects (n=849) ranging in age from neonates (birthweight > 2500g) to adults 23-62y. Those with a history of blood transfusions, major surgery, recent infections, or disease were excluded. Participants were nonanemic (i.e., Hb ≥ 110g/L) with no evidence of iron deficiency. Measurements were performed using an ELISA kit (Orion Diagnostica, Espoo, Finland) that is no longer available (Table 17.31). (Choi et al., 1999).

Table 17.31. Serum transferrin concentrations in various apparently "healthy" nonanemic populations. Data for neonates from cord blood, calculated using non-parametric methods. Data from Choi et al. (1999).

	Age	Mean ± SD (mg/L)	95% CI (m
Neonates	0-5min	4.95 ± 1.24	2.21-6.74
Children			
— Infants	4-24 mo	4.51 ± 1.12	2.15-6.31
— Young children	3-7 y	3.02 ± 0.76	1.47-4.24
Adolescents			
— Middle school	14-16 y	2.86 ± 0.74	1.35-4.19

	Age	Mean ± SD (mg/L)	95% CI (m
— High school	17-19 y	2.09 ± 0.55	1.18-3.23
Adults	23-62 y	2.13 ± 0.51	1.22-3.31

- Finnish Tables of 2.5th and 97.5th percentile reference limits with 95% CI for children 6mos to 16y categorized into four age groups (6mos-4y,n=52; 4-10y, n=60; 10-16y, n=160; > 16y (Suominen et al., 2001). Those with conditions known to affect iron status were excluded. The clinical validity of the reference limits were confirmed for each age group by comparison of stained bone marrow samples with serum sTfR values from 24 children 1.25-16.5y (Suominen et al., 2001). Measurements were performed using an automated immunoturbidimetric assay (IDeAsTfR-IT; Orion Diagnostica) on a Hitachi 917 analyzer.

17.9.3 Cutoff values indicative of risk tissue iron deficiency

value for serum sTfR concentration because of the large assay differences in the measurement of sTfR, as noted earlier. These discrepancies reflect the differences in preparations of transferrin receptor used to raise antibodies and as a standard in the various assays. As a consequence, assay-dependent cutoffs must be applied. There is an urgent need to establish internationally applicable cutoff values for serum sTfR concentrations.

Cutoffs for breastfed infants at age 4, 6, and 9 mos were developed by Domellöf et al. (2002a); these are shown in Table 17.32. No details of the assay used are provided.

Table 17.32. Suggested –2SD cutoff values for iron status variables for infants at 4, 6, and 9mo, based on iron-replete, breastfed infants. MCV, mean cell volume (data from Swedish infants); ZPP, zinc protoporphyrin; TfR, transferrin receptor. From Domellöf et al. (2002a).

Iron status variables	4mo	6mo	9mo
Hb (g/L)	< 105	< 105	< 100
MCV (fL)	< 73	< 71	< 71
ZPP (µmol/mol heme)	> 75	> 75	> 90
Ferritin (µg/L)	< 20	< 9	< 5
sTfR (mg/L)	> 11	> 11	> 11

Cutoff for preschool children 1-5y based on the 97.5th percentile in a U.S. “healthy” reference population was 6.0mg/L (Table 17.30; Mei et al., 2012). The ROCHE Tina-quant sTfR assay was used.

Cutoff for non-pregnant women 15-49y based on the 97.5th percentile in a U.S. “healthy” reference population was 5.3mg/L (Table 17.30) (Mei et al., 2012). The ROCHE Tina-quant sTfR assay was used.

Cutoff for pregnancy of > 4.4mg/L indicative of iron deficiency was used for both U.S. pregnant women from U.S. NHANES 1999-2006 (Mei et al., 2011) and for pregnant women in Kenya (Bah et al., 2017). This cutoff was derived by the manufacturer of the Roche Tina-quant ferritin immunoturbimetric assay (Kolbe-Busch et al., 2002). It is of interest that use of this sTfR cutoff yielded a prevalence estimate for iron deficiency that was comparable to that based on total body iron < 0mg/kg; for more details on total body iron (TBI), see “Body iron” in Section 17.10.1. Several other studies have applied a cutoff of sTfR > 8.5mg/L, said to be indicative of tissue iron deficiency, as determined by quantitative phlebotomy (Skikne et al., 1990), as well as specific for iron deficiency during pregnancy (Carriaga et al., 1991; Akesson et al., 1999; Rusia et al., 1999). These studies have all employed the RAMCO enzyme immunoassay which yields values that are on average 30% higher than those with the ROCHE Tina-quant automated sTfR immunoturbidity assay (Pfeiffer et al., 2007).

17.9.4 Measurement of serum soluble transferrin receptor

Serum is the preferred matrix although some assays can use EDTA-heparin or citrate plasma. Both venous and capillary blood samples can be used but must be refrigerated immediately after collection. Serum samples can be stored for no longer than 2d at room temperature, 7d at 4°C, and 1y at < −20°C. Pfeiffer and Looker (2017) report serum sTfR was stable for > 10y when stored at −70°C. Repeated freezing and thawing is not recommended, although < 3 freeze-thawing cycles do not negatively affect serum sTfR concentrations. Moderate hemolysis does not confound the results. Assays on whole-blood spots dried on filter paper are now possible, although not widely used (Cook et al., 1998; McDade et al., 2002).

In the past, two-site immunoradiometric assays and ELISA were used to measure serum sTfR. Today, measurement of serum TfR in clinical laboratories is carried out via immunoassays using commercial kits on automatic clinical analyzers (e.g., Tina-quant sTfR, Roche Diagnostics) (Kolbe-Busch et al., 2002). However, clinical analyzers are expensive, require a relatively large sample volume (typically > 150µL), and utilize reagents that are relatively costly, with poor comparability across assays. In low income countries, manual ELISA assays available from a few manufacturers are still used, requiring a smaller sample volume (< 50µL), and a relatively inexpensive microplate reader, although their precision is only moderate (Lynch et al., 2018). The RAMCO enzyme immunoassay yields values that are on average 30% higher than those from the Roche Tina-quant sTfR assay (Pfeiffer et al., 2007).

An international sTfR certified reference material is now available from WHO through the United Kingdom National Institute of Biological Standards and Control [sTfR (NIBSC RR 07/202)]. Development of this certified reference material will enable different laboratories to calibrate their own assays, facilitating comparisons across countries (Thorpe et al., 2010).

This page titled 14.9: Serum soluble transferrin receptor (17.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.10: Multiple biomarkers (17.10)

Table 17.33. Relationships between biomarkers and iron status. From Lynch et al. (2018).

Iron status	Sustainable bone marrow iron	Serum ferritin	Transferrin saturation	Eryt
Iron-deficiency anemia	Absent	Low	Low	High
Iron-deficient erythropoiesis	Absent	Low	Low	High
Iron depletion	Absent	Low	Normal	Norr
Normal iron status	Normal	Normal	Normal	Norr
Iron overload	Normal or increased	High	High	Norr

The use of several concurrent biomarkers of iron status provides a more valid assessment of iron status than measurement of any single biomarker. Such an approach minimizes the misclassification that can occur due to the overlapping of normal and abnormal values when a single measure is used.

Moreover, use of multiple biomarkers enables the three stages in the development of iron-deficiency anemia (iron depletion, iron-deficient erythropoiesis, and iron-deficiency anemia) to be characterized. In this way, the severity of iron deficiency can be differentiated more readily. In addition, normal iron status and iron overload can also be identified. The relationship between the biomarkers commonly used (serum ferritin, transferrin saturation, erythrocyte protoporphyrin, serum sTfR, Hb) and the gold standard test — sustainable bone-marrow iron — with iron status is shown in Table 17.33.

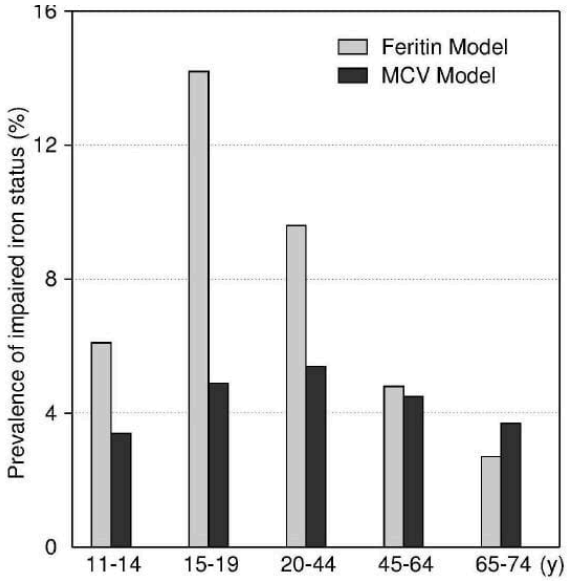


Figure 17.16. prevalence of impaired iron status in subjects of varying ages estimated using the ferritin model and the MCV model: U.S. NHANES II, 1976-1980. Redrawn from Pilch and Senti (1984).

In the past, a combination of three biomarkers of iron status has been used, with abnormal values for at least two of the three bio-

markers indicating iron deficiency. If iron deficiency was concurrent with a low Hb concentration, then individuals were considered to have iron deficiency anemia (Looker et al., 1997). In U.S. NHANES II and U.S. NHANES III(Pilch and Senti, 1984; Dallman et al., 1984; Looker et al., 1997), two models were used to assess the relative prevalence of impaired iron status and anemia in selected groups of the U.S. population. The first (ferritin model) was based on serum ferritin, transferrin saturation, and FEP; whereas the second (MCV model) used MCV rather than ferritin, together with transferrin saturation, and FEP. Figure 17.16 compares the prevalence of impaired iron status by age group in individuals participating in the U.S. NHANES II, 1976-1980 based on the ferritin and MCV models. In most age groups, as expected, the prevalence of impaired iron status was higher using the ferritin model than the MCV model, most notably for the younger age groups.

The cutoffs used by NHANES to identify abnormal values of iron status based on the three biomarkers used in the ferritin model and included in the analysis of selected data from NHANES are shown in Table 17.16. (Looker et al., 1997). These cutoffs to define abnormal values were validated whenever possible in clinically diagnosed adult patients, and therefore are “true cutoffs”, with the exception of those for children and adolescents. Subsequently these same cutoffs were recommended by WHO / UNICEF / UNU in (2001 together with cutoffs for iron overload. Note the differences in the mode of expression for erythrocyte protoporphyrin between Table 17.16 and Table 17.34.

Table 17.34. Cutoff values for erythrocyte protoporphyrin, serum ferritin, and transferrin saturation by stages of iron status and by po group. From Lynch et al. (2018).

	Population group	
	< 5y of age	≥ 5y of age
Erythrocyte protoporphyrin		
Iron overload	Normal	Normal
Normal iron status	Normal	Normal
Iron depletion	Normal	Normal
Iron deficiency with or without anemia	> 70g/dL RBC	> 80µg/dL RBC
	> 2.6µg/g Hb	> 3.0µg/g Hb
	> 61mmol/mol heme	> 70mmol/mol heme
Serum ferritin, µg/L		
Severe risk of iron overload	No cutoff	> 200 (adult males) > 150 (adult females)
Depleted iron stores	< 12	< 15
Transferrin saturation		
Iron overload		> 60-70%
Iron deficiency anemia		< 16%

Increasingly, the inclusion of serum sTfR as one of the laboratory criteria to assess iron status is recommended because it is a relatively sensitive and reliable guide to the degree of functional iron deficiency (i.e., iron-deficient erythropoiesis) (Skikne et al., 1990), and less affected by inflammation than serum ferritin (Rohner et al., 2017).

The NHANES III 2003-2006 survey was the first U.S. NHANES survey to include a combination of biomarkers that included serum sTfR, serum ferritin, and Hb to describe the iron status of a population. The rationale for this combination was that while serum ferritin levels reflect the decline of body iron stores, it does not reflect the severity of the depletion once the iron stores are exhausted. In contrast, serum sTfR concentrations increase after iron stores are exhausted, when they continue to rise with

increasing functional iron deficiency (i.e., in iron-deficient erythropoiesis).

Furthermore, such a combination provides diagnostic information over a wide spectrum of iron status, ranging from replete iron stores to overt iron deficiency anemia. Cutoff values for serum sTfR are not included in Table 17.34 because assay-dependent cutoffs must be applied.

Following the validation of the Total Body Iron (TBI) model (Cook et al., 2003), together with the development of fully automated assays for both serum ferritin and sTfR , U.S. NHANES III replaced the ferritin model with the Total Body Iron (TBI) model to assess the prevalence of iron deficiency, as shown in Table 17.35. See “Body iron” in Section 17.10.1 for more details on the derivation of TBI.

In 2007 WHO developed guidelines on the use of serum sTfR together with ferritin to assess the iron status of populations, and presented a classification of iron deficiency based on these two biomarkers which is shown in Table 17.36.

Table 17.36. The interpretation of low serum ferritin and high transferrin receptor concentrations in population surveys: this classific requires validation in popu lation surveys (WHO, 2020).

[a] *Apply thresholds by age group given in WHO / UNICEF /UNU(2001).*

[b] *Apply thresholds recommended by manufacturer of assay until an international reference standard is available.*

[c] *< 10% for pregnant women.*

[d] *≥ 30% for pregnant women.*

Percentage of serum ferritin values below threshold[a]	Percentage of transferrin receptor values above t
< 20%[c]	< 10%
< 20%[c]	≥ 10%
≥ 20%[d]	≥ 10%
≥ 20%[d]	< 10%

Note that the calculation of TBI was not recommended by WHO (2007). They advised that for prevalence estimates, a single number based on the prevalence from serum ferritin should be used except in settings where inflammation is prevalent when the prevalence based on sTfR is more appropriate. Where possible, when determining prevalence estimates, biomarkers of inflammation (CRP and AGP) should also be assayed, so that both serum ferritin and sTfR concentrations can be adjusted for inflammation, preferably using the BRINDA correction method described in Section 17.7.3.

WHO (2007) have also made recommendations on the use of the most appropriate iron biomarkers to evaluate the impact of interventions to control iron deficiency, and to predict a change in Hb concentrations in response to iron interventions in populations (e.g., supplementation or fortification). Based on data from nine double-blind, randomized controlled trials in which iron was provided as supplements or as food fortified with iron for periods of 4-18mos, serum ferritin and Hb were concluded as the most efficient indicators of population response to iron interventions. Serum ferritin or serum sTfR, although less successful, were considered the best to predict a change in Hb concentration in response to iron interventions. Furthermore, WHO advised that if both serum ferritin and serum sTfR were measured, then body iron stores could be estimated to predict the change in Hb, provided the correct algorithm was applied; for more details see Mei et al. (2005) and WHO (2007).

It is noteworthy, that in settings where iron deficiency is the major cause of anemia, then the Hb concentration is likely to improve more rapidly than serum ferritin following an iron intervention. However, when other factors in addition to iron are the cause of anemia, then an improvement in serum ferritin rather than Hb is more likely. Again, biomarkers of inflammation (CRP and AGP) should also be assayed, where possible, so that serum ferritin and serum sTfR concentrations can be adjusted for inflammation, where necessary.

17.10.1 Use of ratios of sTfR to serum ferritin as indicators of iron status

Increasingly, the ratio of sTfR to serum ferritin is being used as an indicator of the extent of iron deficiency. Serum ferritin, as noted in Section 17.7, is a sensitive indicator until body iron stores (mainly in hepatocytes) are depleted, but once stores are totally exhausted (i.e., with serum ferritin < 12g/L), serum ferritin does not provide information on the severity of the iron deficiency. In

contrast, serum sTfR is a sensitive indicator after body iron stores are exhausted, with concentrations progressively increasing as iron deficiency advances. Hence, in combination, these two biomarkers assess the full range of iron status from severe deficiency to overload. Moreover, when expressed as a ratio, they correct for any plasma volume expansion differences during pregnancy, thus providing a more accurate reflection of iron status at this time. Nevertheless, the practical utility of the sTfR/ ferritin ratios in serum for detecting iron deficiency in the presence of inflammation is still uncertain (Lynch et al., 2018).

Several approaches exist to express and to interpret this ratio, each with different cutoff values to interpret the ratio; these approaches and cutoff values are shown in Table 17.37, and discussed below.

Table 17.37. Approaches to express and interpret the ratio of sTfR to SF. ACD, anemia of chronic disease; ELISA, enzyme linked imm anemia of chronic disease; SF, serum ferritin; sTfR, soluble transferrin receptor. [2] Units of both sTfR and SF in the ratio are µg/L. Da

Ratio	Calculation	sTfR assay used to establish ratio
Total body iron,[2] mg/kg	–[log(sTfR/SF) –2.28229]/0.1207	In-house ELISA that is equivalent to Ramco assa relation to Roche assay
Simple ratio[2]	sTfR/SF	In-house ELISA that is equivalent to Ramco assa relation to Roche assay
sTfR index, mg/L	sTfR/$\log_{10}$SF	First ELISA from R&D Systems; then adopted on Coulter Counter

Simple ratio of sTfR to ferritin(both expressed in µg/L) was calculated in the initial phlebotomy study of Skikne et al.(1990) and is easier to calculate than body iron. Their results confirmed that because of the reciprocal relationship between serum sTfR and ferritin measurements, the ratio of sTfr:SF represented iron status over the entire range in their study. When plotted logarithmically, the ratio increased from < 100 in those with ample iron stores to > 2,000 in those with significant functional iron deficiency. A rise above 500 occurred when stores were fully depleted (i.e., iron stores of 0mg/kg). A major disadvantage of this simple ratio is that its interpretation is assay dependent. The cutoff developed for the simple ratio (i.e., ≤ 500 indicative of ample iron stores; > 500 of depleted iron stores) was established with the Ramco assay, and should only been used with data generated with either the Ramco assay (van den Brock et al., 1998) or an assay equivalent (Grant et al., 2012).

sTfR index, mg/L is calculated as the ratio of sTfR to the base-10 logarithm of serum ferritin (Table 17.37) and was developed initially to identify individuals with depleted iron stores, defined by a complete absence of stainable iron in the bone marrow (Punnonen et al., 1997). Use of the log of serum ferritin in this ratio decreases the influence of the acute phase response on the ferritin component of the ratio (WHO 2007). Cutoff values have been developed (expressed in mg/L) to distinguish between the anemia of chronic disease (< 1mg/L) and iron deficiency anemia (> 2mg/L), or both conditions (> 2mg/L)(Weiss and Goodnough, 2005) (Table 17.37). The sTfR index was later adopted for use with the assay of sTfR and ferritin simultaneously by the Access immunoassay systems (Beckman Coulter Counter), when a cutoff of > 1.03mg/L was proposed for iron deficiency anemia or a combination of iron deficiency anemia and the ACD. However, the interpretation of the sTfR index is assay dependent so this proposed cutoff should not be applied when alternative assays not comparable to the Beckman Coulter Counter are used (Pfeiffer and Looker, 2017). Hence, standardization of the various sTfR assays is needed before this ratio is widely used. To convert sTfR measurements when expressed as nmol/L to mg/L, multiply nmol/L by a factor of 0.0738 (Skikne et al., 2011).

Body iron was the original term used by Skikne et al., (1990), although others have used body iron stores or total body iron (TBI). Measurements of body iron were initially based on a study in which serial measurements of serum ferritin and sTfR were obtained during repeated phlebotomies in 14 healthy Caucasian adults (6M; 8F, 24-46y) over 6-22wks; see Skikne et al. (1990) and Cook et al. (2003) for further details. The investigators demonstrated a close linear relationship between the logarithm of the ratio of sTfR/serum ferritin and total body iron stores expressed as mg per kg body weight. The formula for this relationship (based on both sTfR and serum ferritin (SF) expressed as µg/L) is the following:

$$\text{Body iron (mg/kg)} = \frac{[\log_{10}(\text{sTfR/SF}) - 2.8229]}{0.1207}$$

Box 17.5: Advantages of using total body iron (TBI) (as µg/kg) as a measure of iron status

- Independent from Hb, thereby shifting the focus from screening and prevention of anemia to iron deficiency
- Fewer laboratory measurements required compared to the ferritin model

- Provides an estimation of entire distribution of iron stores based on body weight
- Assay methods for the calculation of TBI are readily automated
- Adequate agreement between prevalence of iron deficiency based on TBI and ferritin model in preschool children and women of childbearing age in the U.S. NHANES 2003-2006
- Positive values of TBI represent iron stores and negative values represent tissue iron deficiency
- Suggested cutoff for defining iron deficiency is < 0mg/kg. However, no cutoffs exist for iron replete and excess.

Modified from Gupta et al. (2017).

This formula was validated by Cook et al. (2003) using data from three published studies. Note this model is based on serum sTfR concentrations using an in-house ELISA assay (Flowers et al., 1986), shown later by Pfeiffer et al. (2007) to be equivalent to the Ramco sTfR assay. The Ramco assay was later reported to measure ≈ 30% higher than the fully automated sTfR immunoturbimetric assay developed by Roche Diagnostics and used in the U.S. NHANES III survey to measure body iron (Cogswell et al., 2009). Consequently, Pfeiffer et al. (2007) developed a regression equation to convert serum sTfR concentrations assayed using the fully automated ROCHE assay to Flower's sTfR concentrations:

$$\text{Flower's sTfR} = 1.5 \times \text{Roche'sTfR} + 0.35(\text{mg/L})$$

Note this body iron equation cannot be used directly when other sTfR assays are used. Factors that led to the decision to apply the body iron model to NHANES data are itemized in Box 17.5.

Figure 17.17 depicts TBI, expressed as the prevalence of low iron stores for the first, second, and third trimester in U.S. pregnant women from U.S. NHANES 1990-2006 survey. In this study, the sTfR concentrations for each trimester were adjusted to be comparable with values produced by the Flowers sTfR assay used in the development of the TBI model, by using the regression equation shown above. Results confirmed that iron deficiency based on TBI < 0mg/kg was more prevalent in pregnant women in the second or third trimester, and in non-Hispanic black pregnant women (data not shown). Moreover, the prevalence and pattern, by trimester and by race / ethnic group (data not shown) of iron deficiency based on sTfR and TBI were similar.

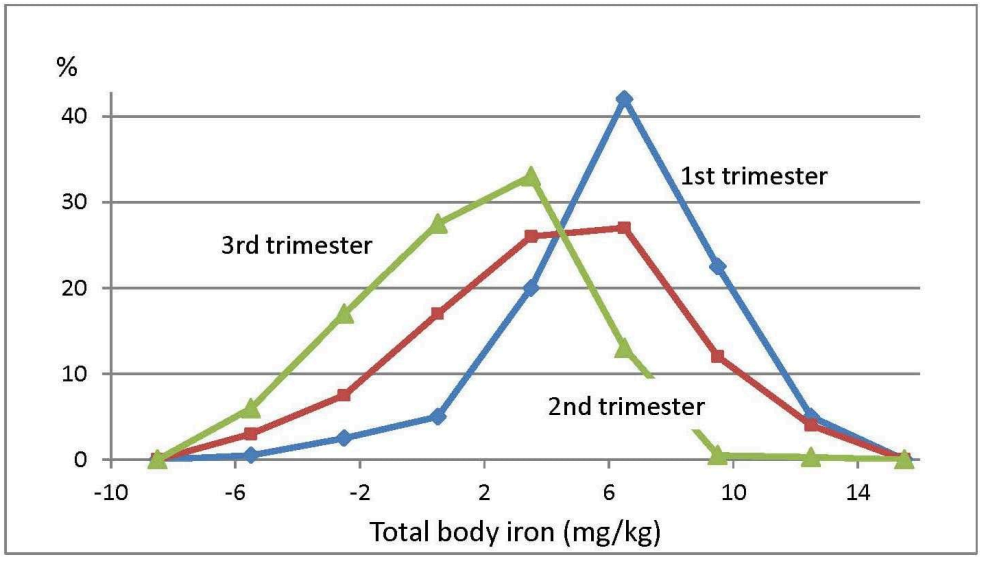


Figure 17.17. Distribution of total body iron (calculated from serum ferritin and sTfR concentrations) in U.S. pregnant women at the 1st, 2nd, and 3rd trimester. Data from the U.S. NHANES 1999-2006. Redrawn from Mei et al. (2011).

Note that these results did not take inflammation and infection into account, even though pregnancy is considered an inflammatory state. However, after adjusting the prevalence of iron deficiency based on TBI for inflammation by excluding those with elevated CRP concentrations (i.e., > 5mg/L), the prevalence of iron deficiency after exclusion was not substantially different, suggesting that the effect of inflammation on the prevalence of iron deficiency in these U.S. women during pregnancy was small

(Table 17.38).

Nevertheless, several uncertainties remain when using TBI to define iron deficiency. As an example, whether the equation for estimating total body iron is appropriate for young children and pregnant women is unclear. The equation was validated only in non-pregnant adults (Cook et al., 2003), and the relationship between TBI and serum ferritin and sTfR concentrations may differ in these life-stage groups. Furthermore, the changes in serum ferritin and sTfR that occur during more gradual reduction in body iron stores may not resemble those observed during the relatively rapid reduction in body iron stores with phlebotomy (Mei et al., 2017).

Table 17.38. Unadjusted prevalence of iron deficiency (ID percentage: 95%CI) on the basis of high serum soluble transferrin recept pregnant women in the National Health and Nutrition Examination Survey (NHANES), 1999-2006. Data from Mei et al. (2011).

Trimester	n	High sTfR (> 4.4mg/L)
First	189	5.8 (2.4, 9.2)
Second	416	10.6 (6.3, 14.9)
Third	384	27.9 (20.6, 35.3)

There is also some confusion over the term TBI because it is not a measure of the quantity of iron in the body of an individual. Instead, total-body iron provides a quantitative estimate of the size of the body store when iron is present in the store (i.e., values > 0mg/kg). In the case of iron deficiency (i.e., values < 0mg/kg) , TBI provides an estimate of the quantity of iron that would be needed to replace the functional deficit (Pfeiffer and Looker, 2017). In light of the confusion, an alternative term — body iron index — has been proposed.

In addition, there is some concern over the assumption that a body iron index value of "0" is in fact the criterion for defining iron deficiency in view of inconsistencies in the reported prevalence of iron deficiency estimated by serum / plasma ferritin, sTfR, and TBI (Engle-Stone et al., 2013; Cogswell et al., 2009). This has prompted the suggestion that a higher cutoff such as 2mg/kg may be more appropriate than TBI of < 0mg/kg to indicate iron deficiency; for more discussion of these concerns, see Lynch et al. (2018).

In the future, the development of low cost ELISA assays capable of measuring several iron and inflammation status biomarkers simultaneously from a small volume of serum / plasma will assist in improving the reliability of the diagnosis of iron deficiency. This is especially needed in settings where a high burden of infectious diseases, genetic Hb disorders, and micronutrient deficiencies have often confounded the interpretation of the data.

This page titled 14.10: Multiple biomarkers (17.10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

14.11: Emerging iron status indicators (17.11)

Emerging indicators, currently mainly used in a research setting, include hepcidin, non-transferrin-bound iron, and some reticulocyte indices. More work is required to understand the diagnostic value of these emerging indicators before they become more widely available for public health or clinical use.

17.11.1 Hepcidin

Hepcidin is a protein encoded by the HAMP gene. It plays a major role in controlling physiological iron homeostasis, as noted earlier. Hepcidin binds and subsequently causes degradation of the iron-exporter ferroportin, thereby inhibiting both absorption of dietary iron from the gastrointestinal tract and the release of iron from body stores (Section 17.1.2). In conditions of iron depletion or increased iron demand, hepatic synthesis of hepcidin is reduced, thus facilitating iron absorption and mobilization from body stores via active ferroportin. In contrast when iron stores are replete (or in renal impairment when excretion of hepcidin is suppressed), hepcidin levels are increased which block the release of iron into the circulation. Instead, iron is lost with the cell when it is shed from the villus. Hepcidin production by the liver is also regulated by inflammation and erythropoietic activity.

Hepcidin may be useful in the clinical diagnosis of iron-refractory anemia (i.e. the type of iron deficiency anemia that typically does not improve with oral iron treatment), in differentiating between iron deficiency anemia and anemia of chronic disease and in patients with iron overload syndromes. As hepcidin is a determinant of dietary iron absorption, hepcidin may also provide guidance on safe iron supplementation in countries with a high infection burden (van der Vorm et al., 2016; Pfeiffer and Looker, 2017). Studies have also investigated the diagnostic potential of serum hepcidin as an index of iron deficiency in pregnancy (Bah et al., 2017). Nevertheless, whether hepcidin assays provide an advantage over serum ferritin and the other more common methods for assessing nutritional iron status, particularly when iron status is replete, remains uncertain (Pfeiffer and Looker, 2017; Lynch et al., 2018).

Inflammation raises levels of pro-inflammatory cytokines in the systemic circulation, which in turn increase serum hepcidin levels, an acute-phase protein, just like serum ferritin. Elevated hepcidin levels reduce both iron intestinal absorption and the release of iron from body stores, thus decreasing the amount of circulating serum iron and thereby limiting erythropoiesis. Hence, the risk of iron deficiency increases, even though body iron stores are normal.

Obesity is associated with low grade inflammation and thus elevated hepcidin levels. Results presented in Table 17.39

Table 17.39. Inflammatory markers and iron biomarkers in normal weight healthy young women (n=22) and obese women (n=25).
WBC, White Blood Cell Count; IL-6, interleukin-6; CRP, C-reactive protein; MCV, mean corpuscular volume (mean cell volume); TIBC, total iron-binding capacity; TSAT, transferrin saturation.
[1] Data available for 40 women (18 normal weight vs. 22 obese);
*[2] Data available for 42 women (20 normal weight vs. 22 obese); * Geometric mean (95% CI).*
Abstracted from Aguree and Reddy (2021).

Variable	Normal Weight Women (BMI: 18.5-24.9 kg/m^2) Mean ± SD	Obese Women (BMI: >29.9 kg/m^2) Mean ± SD	p Value
Inflammatory markers			
WBC (×10^{3}µL)	5.6 ± 1.6	7.7 ± 2.0	< 0.001
IL-6 (pg/mL) [1*]	1.46 [1.13, 1.89]	2.16 [1.86, 2.51]	0.003
CRP (mg/L) [1*]	8.2 [3.1, 21.8]	69.9 [41.1, 118.9]	< 0.001
Iron biomarkers			
Hb (g/dL)	13.6 ± 1.1	13.3 ± 1.1	0.166
MCV (fL)	86.3 ± 4.5	82.4 ± 8.9	0.030

Variable	Normal Weight Women (BMI: 18.5-24.9 kg/m^2) Mean ± SD	Obese Women (BMI: >29.9 kg/m^2) Mean ± SD	p Value
Serum iron (µg/dL) [2]	112.0 ± 41.4	92.4 ± 33.9	0.044
TIBC (µg/dL) [2]	394.4 ± 61.5	371.3 ± 50.4	0.086
TSAT (%) [2]	29.7 ± 13.3	25.2 ± 9.1	0.094
Ferritin (ng/mL) [1*]	34.0 [21.0, 55.1]	37.7 [26.9, 52.9]	0.355
Hepcidin (ng/mL) [1]	6.21 [4.39, 8.77]	11.21 [7.04, 17.83]	0.024

compare inflammatory and iron status biomarkers in a cross-sectional study of normal weight and obese healthy U.S. women. The data highlight the higher mean serum concentrations of CRP, interleukin-6, and hepcidin but lower levels of MCV and serum iron (but not ferritin) in the obese women compared to their normal weight counterparts (Aguree and Reddy, 2021).

Enhanced erythropoiesis suppresses hepcidin because of the increased iron demand. Consequently, iron absorption is increased and iron is mobilized from body stores. Numerous conditions are associated with an increase in erythroid proliferation; some examples are given in Table 17.28. Of concern are individuals with severe transfusion-dependent HbE β-thalassemia and β-thalassemia carriers (but not HbE carriers) because the increased uptake of iron resulting from severe and mildly suppressed hepcidin, respectively, may increase the risk of iron overload. As an example, Table 17.40 highlights the suppression

*Table 17.40. Demographic, iron, erythropoietic, and inflammatory indices for patients with HbE β-thalassemia stratified by severe and moderate phenotype, NTBI, Non-transferrin-bound iron. Values are *mean (range) or **geometric mean (range). †significant difference between moderate and severe patients (P < 0 .05). ‡significant difference between controls and severe patients, and controls and moderate patients. Data from Jones et al. (2015).*

Parameter	Severe (n=28; 17 female)	Moderate (n=41; 25 female)	*P*	Local controls (n=25)
Hb g/dL*	5.8 (4.4-7.6)	6.4 (4.9-8.3)	.1397	15.0 (14.5, 15.5)‡
Ferritin mg/L**	1356 (328, 9790)	732 (143, 3260)	.0021†	48.2 (38, 61.2)‡
Liver iron mg/g dwt**	11.3 (1, 54.2)	5.30 (0.7, 29.0)	.0006†	
NTBI µM/L*	5.16 (–1.67, 10.27)	4.36 (–2.67, 23.75)	.4456	
Hepcidin ng/mL**	2.24 (0.1, 51.8)	1.95 (0.1, 36.3)	.7441	28.23 (18.83, 42.3)‡
sTfR nM/L**	6.51 (1, 35.3)	8.87 (1.4, 35.3)	.1472	
CRP mg/L**	1.66 (0.23, 34.23)	1.58 (0.20, 13.01)	.8433	
Labile plasma iron µM/L*	3.60 (0.1, 7.7)	4.29 (–1.07, 13.61)	.9320	
Transferrin SAT, %*	97.3 (48.3, 100.0)	86.6 (31.9, 100.0)	.0319†	

of hepcidin in Sri Lankan patients with HbE β-thalassemia classified as severe or moderate clinical phenotypes. Patients in both groups had lower concentrations of serum hepcidin and Hb and higher serum ferritin concentrations when compared to the values for 25 iron-replete, non-thalassemic Sri Lankan male controls.

Even among Sri Lankan schoolchildren, hepcidin was mildly suppressed among those with increased erythropoiesis associated with β-thalassemia trait (but not with HbE) compared to controls, suggesting an enhanced propensity to accumulate iron (Jones et al., 2015). Hence, in Asian countries where β-thalassemia trait syndromes and thalassemia carriers are prevalent, risk of causing an inadvertent burden of iron overload should be considered when planning public health iron interventions to improve the iron status of high risk population groups (Jones et al., 2015), as noted earlier.

Malaria induces an increase in hepcidin concentrations, likely as a result of suppressed erythropoietic activity, inflammation and perhaps additionally by a direct stimulatory effect of malaria parasites and their products (de Mast et al., 2009). Hepcidin concentrations are elevated even when the malaria is asymptomatic (i.e., presence of parasitemia in the absence of fever or malaria-related symptoms) and in the absence of a marked acute phase response, as shown in Figure 17.18. Although an effect of mild inflammation cannot be excluded, nevertheless, other mechanisms might also contribute to elevated hepcidin concentrations such as the existence of an IL-6-independent pathway in malaria.

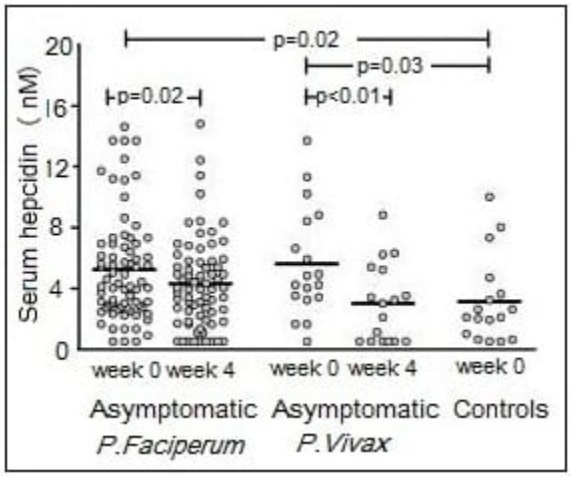


Figure 17.18. Serum concentrations of hepcidin in children with asymptomatic *P. falciparum* parasitemia (n=73), asymptomatic *P. vivax*parasitemia (n=18) and controls (n=17). Data depicted are before and 4 weeks after the start of antimalarial treatment; the line represents the mean. P values were determined using the Student's t-test and paired t-test. Redrawn from de Mast et al. (2009).

The results shown in Figure 17.18 are based on a group of Indonesian children (5-15y) (n=1197) screened by microscopy for the presence of parasitemia, Hb, serum hepcidin, indices of iron status (ferritin, sTfR, mean cell volume) and inflammation (CRP) at baseline and 4 weeks after antimalarial treatment.

It is of interest that the presence of even mildly elevated hepcidin concentrations appear to induce functional iron deficiency in these Indonesian children with asymptomatic parasitemia, as indicated by lower values of Hb, MCV, serum iron and transferrin saturation. These findings suggest that in the presence of asymptomatic malaria, iron therapy for the treatment of malaria anemia may be less effective because absorption of iron is compromised by the high hepcidin concentrations. Moreover, the iron therapy may be even hazardous, increasing malaria-associated morbidity and mortality (de Mast et al., 2010).

Pregnancy suppresses hepcidin by a mechanism that is unclear (Koenig et al., 2014). During the first trimester, serum hepcidin concentrations are within the reference range for nonpregnant women, but decrease in the second trimester to very low levels, where they remain in the third trimester, increasing again immediately after delivery and thereafter (van Santen et al., 2013; Bah et al., 2017; Fisher and Nemeth, 2017) (Figure 17.19).

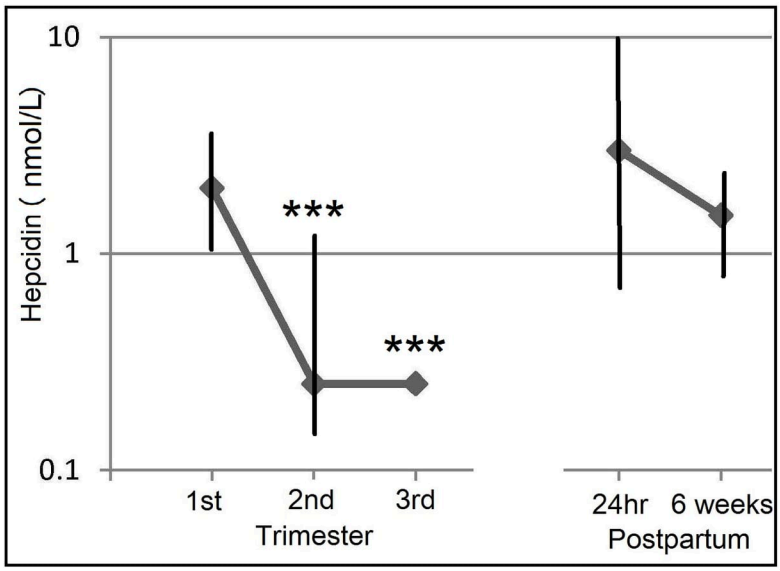


Figure 17.19. Median (IQR) serum hepcidin concentrations in 31 women during pregnancy and postpartum. ***Compared with first-trimester values, P < 0.0001. Modified from Fisher and Nemeth (2017).

The fall in hepcidin appears to occur in advance of the onset of low iron stores in pregnancy, as shown from the data for Gambian women presented in Table 17.29. Here, hepcidin, indices of iron stores, erythropoiesis, and inflammation were all measured at 14, 20 and 30wks gestation(Bah et al., 2017). This suggests that maternal iron deficiency is not solely responsible for the sharp decline. Even mothers with replete iron stores have low hepcidin concentrations at delivery, suggesting that maternal hepcidin may be actively suppressed during pregnancy (Fisher and Nemeth, 2017).

However, in pregnancies complicated by infection or inflammation (including obesity), maternal hepcidin concentrations are elevated compared to healthy controls. Hence, in these conditions maternal and fetal iron bioavailability could be compromised in pregnancy, limiting the amount of iron presented for uptake by the placenta and for transfer to the fetus (Koenig et al., 2014).

Interpretive criteria for hepcidin concentrations are difficult to define because concentrations vary widely across assays, as for serum sTfR (Kroot et al., 2009). As a consequence, no usable international reference ranges and reference limits for hepcidin for specified life-stage groups exist at the present time.

Measurement of serum hepcidin with good reproducibility in both plasma and urine is possible using mass spectrometry or immunoassays, although inter-assay variability is large, with no certified reference material currently widely available (Pfeiffer and Looker, 2017). Recently an international calibrator for hepcidin with long-term stability has been developed (available at HepcidinAnalysis.com) which will greatly improve the equivalence between hepcidin measurement procedures in the future; see Van der Vorm et al. (2016); Diepevenn et al. (2019); and Aune et al., (2020) for further details.

Measurement of hepcidin in urine may provide a less-invasive screening for iron status and warrants more investigation. Concentrations are elevated in anemic children with febrile malaria.

17.11.2 Novel red blood cell indices

Novel red blood cell indices are generated by specific models of hematology automated cell counters and include among others, reticulocyte mean Hb content (RHcc) and mean reticulocyte volume (MRV). However, their interpretation is complicated by the different techniques used by the manufacturer and the lack of systematic studies to determine their usefulness to evaluate nutritional anemias.

Reticulocyte Hb content (RHcc) reflects directly the synthesis of Hb in bone marrow precursors and is a measure of the adequacy of the availability of iron. It can be used to differentiate iron deficiency from other causes of anemia. Reticulocyte Hb content is sometimes used in clinical settings to establish iron status in young children with anemia, when other common iron biomarker assays may not be available; low values provide an early marker of iron deficient erythropoiesis. The measurement of

RHcc is not useful in individuals with α and β-thalassemias (including individuals who are heterozygous) when the reticulocyte Hb count is reduced independently of iron stores.

Reticulocyte Hb content can also be used to measure the early response to intravenous iron therapy because it increases significantly after only 48-72h (Buttarello, 2016). The sensitivity and specificity of RHcc for diagnosing iron deficiency is only moderate, with no recommended cutoffs to define iron deficiency. However, RHcc is less affected by inflammation than serum iron, transferrin saturation, and ferritin, although it is influenced by any conditions that cause iron restricted erythropoiesis (Gelaw et al., 2019).

Reticulocyte volume is also an index that can be measured by some newer hematology analyzers. Values for reticulocyte volume decrease rapidly with the development of iron-deficient erythropoiesis and increase rapidly in individuals with depleted iron stores or nutritional macrocytosis after therapy with iron, or vitamin B_{12} and/or folic acid, respectively. However, numeric results depend on the analyzer manufacturer, making comparisons difficult (Buttarello, 2016).

17.11.3 Non-transferrin bound iron

Iron in plasma is tightly bound to transferrin, the iron transport protein that delivers iron throughout the body, as noted earlier (Section 17.6). Cells take up transferrin-bound iron in proportion to the number of cell-surface transferrin receptors (TfR1). When cellular iron levels are sufficient, the uptake of transferrin decreases to limit further iron assimilation and prevent excessive iron accumulation; usually only 20-30% of transferrin is saturated with iron. However, in iron overload conditions, such as hereditary hemochromatosis and thalassemia major, transferrin is fully saturated and non-transferrin-bound iron (NTBI) appears in serum. Note the high concentrations of NTBI in the patients with HbE β-thalassemia in Table 17.40 , which were inversely associated with hepcidin. This trend may reflect the increased release of iron into the plasma that occurs when suppressed hepcidin levels permit enhanced iron absorption and mobilization. The NTBI also correlated with ferritin and liver iron concentrations, suggesting that a high level of NTBI is associated with hepatic iron accumulation (Jones et al., 2015).

There is no known regulatory mechanism for NTBI uptake so NTBI can enter cells readily, where it produces free radicles resulting in cellular and organ damage. Circulating NTBI may also cause an increase in bacterial-pathogenic infections owing to the free ion being utilized by the parasite. Concern has been raised over the rapid absorption of iron supplements given in non-physiological amounts, and their potential for exceeding the capacity for transferrin to bind the circulating iron, resulting in the formation of NTBI concentrations (Prentice et al., 2016).

The measurement of serum NTBI is fraught with problems, often with a 40-fold variation difference in methods, in part due to the heterogeneous nature of the chemical forms of NTBI, as well as large analytical variation. A new automated NTBI assay has been developed which looks promising in terms of reproducibility and comparability with HPLC, one of the most reliable methods to quantify NTBI. Nevertheless, more research is required to define the most relevant forms of NTBI, and understand their clinical importance (Pfeiffer and Looker, 2017).

This page titled 14.11: Emerging iron status indicators (17.11) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

15: VITAMIN A (CHAPTER 18A)

15.1: Introduction (18a.1)
15.2: Indices of vitamin A status (18a.2)

This page titled 15: Vitamin A (Chapter 18a) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

15.1: Introduction (18a.1)

Vitamin A is a generic term for all retinoids that qualitatively exhibit the biological activity of all-*trans*retinol. The various biologically active forms of vitamin A are shown in Figure 18a.1. Certain carotenoids have provitamin A activity. Of these, α-carotene, β-carotene, and β-cryptoxanthin are the most widely studied. β-carotene is unique because it can theoretically yield two molecules of retinal, which are subsequently reduced to retinol.

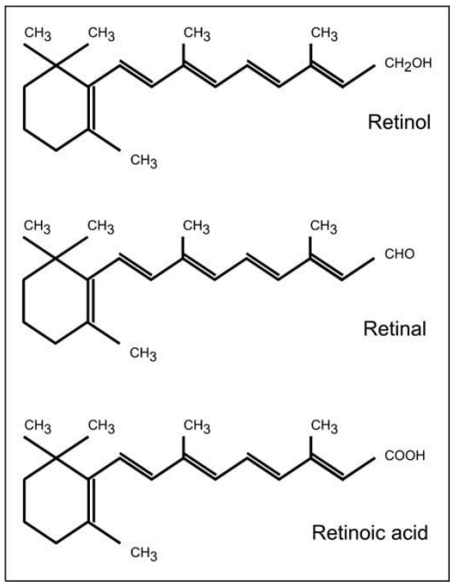


Figure 18a.1: The various forms of vitamin A: retinol, retinal, and retinoic acid.

18a.1.1 Functions of vitamin A

Vitamin A has a clearly defined role in vision: when retinal tissue is deprived of vitamin A, rod and cone function is impaired. However, only the biochemical role of the 11-*cis* retinaldehyde form of vitamin A, in the visual process in rod cells, has been studied in detail. Vitamin A is required for the integrity of epithelial cells throughout the body. The regulatory action of retinoic acid at the level of the gene has an important role in growth, embryonic development, and maintenance of immune function. Both cell-mediated immunity and systemic and mucosal humoral immunity are affected. A review by Tanumihardjo et al. (2016) outlines the various functions and biomarkers of vitamin A.

18a.1.2 Vitamin A deficiency in humans

Early signs of vitamin A deficiency in humans include growth failure, loss of appetite, and impaired immune response with lowered resistance to infection. Xerophthalmia are the sequelae associated with the eye during extreme vitamin A deficiency. Night blindness develops when liver reserves of vitamin A are nearly exhausted. If not corrected, ocular lesions such as conjunctival xerosis, Bitot's spots, keratomalacia, and blindness may occur. In the past, conjunctival impression cytology (CIC) was used to detect early physiological changes characteristic of vitamin A deficiency. Such changes include both the progressive loss of goblet cells in the conjunctiva and the appearance of enlarged, partially keratinized, epithelial cells (Wittepenn et al., 1986; Natadisastra et al., 1988). Filter paper is used to collect the cells and then they are stained. Results have been inconsistent and dependent in part on the severity of the vitamin A deficiency state (Table 18.7) and on which measures are used. Ocular infections may confound the interpretation (Carlier et al., 1992). Due to relatively low sensitivity and specificity, WHO recommends combining this test with other indicators of vitamin A status (WHO, 1996). Currently the method is not being used in human studies.

Vitamin A deficiency still occurs in the world, but major strides have been made in the eradication of blindness through public

health interventions, such as high dose capsule distribution and food fortification with vitamin A. The definition for vitamin A deficiency includes both clinical manifestations-xerophthalmia, anemia, growth retardation, increased infectious morbidity and mortality-as well as the following functional consequences: impaired iron mobilization, disturbed cellular differentiation, and depressed immune response (Sommer & Davidson, 2002).

Severe deficiencies of certain other nutrients may also simulate vitamin A deficiency. Examples include zinc (Christian and West, 1998) and protein-energy malnutrition (Russell et al., 1983); details are given in Section 18.2.1.

Vitamin A deficiency may occur secondary to some disease states, including cystic fibrosis, severe intestinal and liver diseases, and some severe defects in lipid absorption (e.g., cholestasis). In developed countries, the prevalence of frank nutritional deficiency of vitamin A is low. In the U.S. National Health and Nutrition Examination Survey (NHANES III, 1988-1994), for example, the prevalence of low serum retinol concentrations (<0.70µmol/L) was less than 2% in all age, sex, race or ethnic strata (Ballew et al., 2001). The World Health Organization provides guidance on the use of serum retinol concentrations to evaluate population vitamin A status (WHO, 2011).

18a.1.3 Food sources and dietary intakes

Preformed vitamin A is found only in foods of animal origin: fish-liver oils, liver, butterfat, and egg yolk are the major dietary sources. Muscle meats are poor sources of preformed vitamin A. Plant sources, such as nuts, grains, and vegetable oils, have no preformed vitamin A.

Provitamin A carotenoids are found in both plant and animal products, but in low- income countries the main food sources are yellow and orange-colored fruits (West, 2000) and dark-green leafy vegetables. Red palm oil, and certain indigenous plants such as palm fruits (buriti) in Latin America, and the fruit termed "gac" in Vietnam, are unusually rich sources of provitamin A carotenoids (FAO/WHO, 2002).

Provitamin A carotenoids, when derived from ripe yellow- and orange-colored fruits and cooked yellow tubers (e.g., sweet potatoes), appear to be more efficiently converted to retinol than when derived from dark green leafy vegetables (IOM, 2001; West et al., 2002). Processing methods and the food matrix also affect the bioavailability of provitamin A carotenoids (Torronen et al., 1996; Rock et al., 1998; van het Hof et al., 1988).

In more affluent countries such as Canada, the United States, and the United Kingdom, the major sources of preformed vitamin A in the diet are liver, milk, and milk products, followed by fish in the United States and Canada (IOM, 2001) and fat spreads (e.g., fortified margarine) in the United Kingdom(Gregory et al., 1990). The major contributors of provitamin A carotenoids are generally vegetables (Gregory et al., 1990; Chug-Ahuja et al., 1993). Of the provitamin A carotenoids, β-carotene followed by β-cryptoxanthin are the most important. Other carotenoids with vitamin A activity include α-carotene, lutein, lycopene, and zeaxanthin.

Currently two conversion factors are used for calculating the amount of vitamin A activity in foods from provitamin A carotenoids, although the values applied differ across agencies. FAO/WHO (2002) still maintain the use of 1µg retinol equals 6µg of β-carotene and 12µg of other provitamin A carotenoids (mainly α-carotene and β-cryptoxanthin). These same carotenoid / equivalency ratios have also been adopted by the European Food Safety Authority (EFSA, 2017). Furthermore, these two agencies express the substances with vitamin A activity as retinol equivalents (RE), whether they are preformed vitamin A (mainly retinol and retinyl esters) in foods of animal origin or provitamin A carotenoids.

The U.S. Food and Nutrition Board, however, concluded that the bioavailability of provitamin A β-carotene from plant sources is 12µg to 1µg retinol and 24µg to 1µg for other provitamin A carotenoids for healthy individuals. For a detailed justification of these conversion factors, see IOM (2001). The U.S has also adopted the term retinol activity equivalents (RAE) for use when calculating the total amount of vitamin A in mixed dishes or diets. If the IOM (2001) conversion factors are adopted, the vitamin A activity in a foodstuff, expressed as a retinol activity equivalency, can be calculated from the following equation:

$$\begin{aligned}\mathrm{RAE}(\mu g) = &\ \text{retinol } (\mu g) + (\beta\text{-carotene } (\mu g)/12.0) \\ &+(\text{ other provitamin A carotenoids } (\mu g)/24.0)\end{aligned}$$

Such inconsistencies in the specific carotenoids/retinol equivalency ratios applied exacerbate problems when comparing vitamin A values among food composition databases and, in turn, vitamin A intakes across countries. For example, vitamin A intakes

calculated from some food composition data may be higher if the lower bioconversion factors for provitamin A carotenoids recommended by FAO/WHO and EFSA were used, rather than the higher bioconversion facturs adopted by the United States (IOM, 2001).

Some older food composition tables continue to express vitamin A in terms of international units (IU). Use of these older units is no longer appropriate for assessing dietary adequacy of vitamin A and should be discontinued (FAO/WHO, 2002). For more discussion of the confusion that may arise when assessing dietary vitamin A intakes, see Melse-Boonstra et al. (2017).

18a.1.4 Effects of high intakes

Suggestions that vitamin A and its carotenoid precursors are cancer-preventive agents led to increased consumption of large doses of vitamin A. This is a serious health hazard, particularly during pregnancy: hypervitaminosis A has been associated with birth defects (Rothman et al., 1995; Azais-Braesco and Pascal, 2000). Clinical manifestations of vitamin A toxicity include a pseudo brain tumor, skeletal pain, desquamating dermatitis, and hepatic inflammation (Frame et al., 1974; Russell, 2000). Concomitant consumption of ethanol appears to enhance the toxicity of vitamin A (Leo and Lieber, 1999).

A U.S. Tolerable Upper Intake Level (UL) has not been set for β-carotene or carotenoids (IOM, 2000), although β-carotene supplements are not advised for the general population. For preformed vitamin A, the U.S. UL varies according to life-stage group, ranging from 600µg/d for infants to 2,800µg/d for adolescents. For nonpregnant, pregnant, and lactating women, the UL is 3000µg/d (IOM, 2001).

The U.S. UL is not applicable to vitamin A-deficient populations who should receive vitamin A prophylactically. Approximately 80 countries are using vitamin A supplementation (Wirth et al., 2017). The World Health Organization (2011) recommend routine high-dose vitamin A supplementation in developing countries for children between the ages of 6 to 59mo. For infants 6-11mo, 100,000 IU should be given as a single dose every 4-6mo. Children aged 12mo and older should receive 200,000 IU as a single dose every 4-6mo. No other age groups are recommended for high dose supplementation by the World Health Organization.

This page titled 15.1: Introduction (18a.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

15.2: Indices of vitamin A status (18a.2)

Most of the vitamin A in the body is stored in the form of retinyl ester in the liver. Therefore, a measure of liver vitamin A stores is the best index of vitamin A status. The definition of vitamin A deficiency is when liver stores of retinol are below 0.1µmol/g (Tanumihardjo, 2021). Vitamin A is not uniformly distributed in the liver (Olson et al., 1979) and can vary by 8-13% among the liver lobes in humans (Olsen et al., 2018). Furthermore, liver biopsies are impractical in population studies. Instead, serum retinol concentrations are often determined. However, the serum or plasma contains only about 1% of the total body reserve of vitamin A, and concentrations do not reflect body stores until they are severely depleted. Consequently, it is best to use other biochemical and physiological functional tests of vitamin A status in combination with serum retinol concentrations. The most used biomarkers of vitamin A status are shown in Figure 18a.2. Details of tests currently used are summarized in a review article (Tanumihardjo et al., 2016) and are discussed briefly in the following sections.

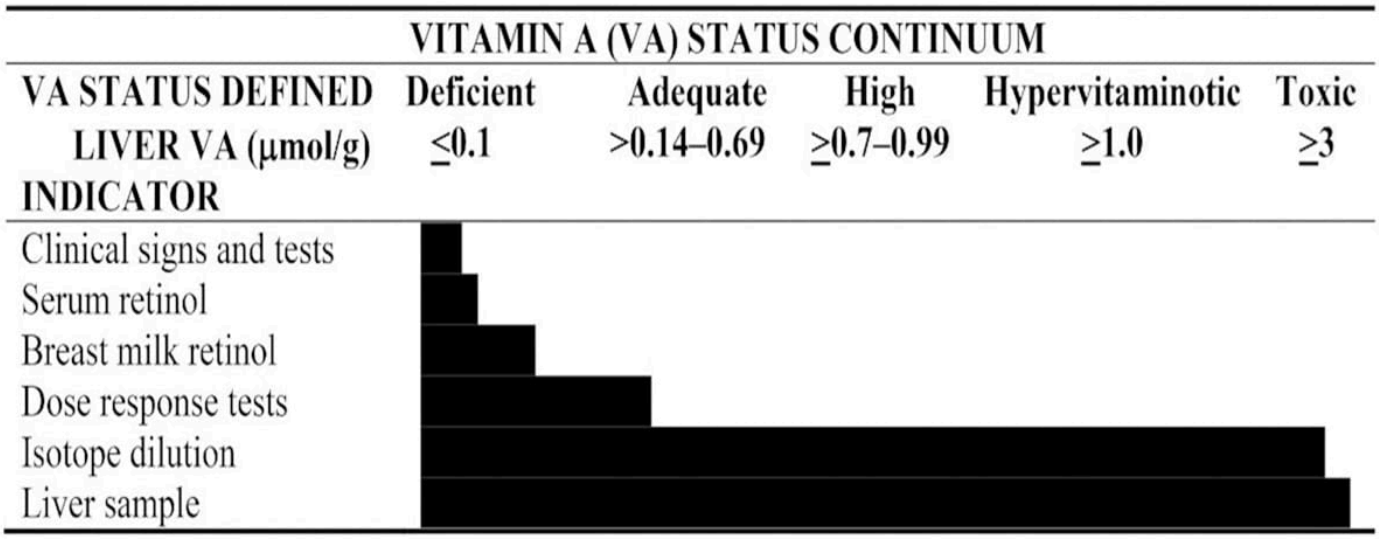


Figure 18a.2: Vitamin A status indicators as they relate to liver vitamin A concentrations, which is considered the reference standard. The length of the black bar indicates the working optimal range (Tanumihardjo, 2020).

18a.2.1 Serum retinol concentrations

Vitamin A in the plasma circulates largely in a 1:1 complex of retinol and retinol-binding protein (RBP). The remainder is in the form of retinyl ester and very small amounts of retinoic acid and other metabolites (Olson, 1984). Retinyl esters are elevated post-prandially after a meal containing preformed or provitamin A. If retinyl esters are elevated in the fasted state, this is an indication of vitamin A toxicity (Tanumihardjo et al., 2016). In humans, this occurs at a liver vitamin A concentration of 3µmol/g liver (Olsen et al., 2018).

Serum retinol concentrations are a common indicator for population studies; however, they only reflect vitamin A status when liver vitamin A stores are severely depleted or excessively high, in which case retinyl esters are also in circulation. When liver vitamin A concentrations are low or adequate, serum retinol concentrations are homeostatically controlled, and levels remain relatively constant and do not reflect total body reserves of vitamin A (Olson, 1984). Hence, it is not surprising that in populations from developed countries, such as the United States, where the liver vitamin A concentrations are generally adequate, positive relationships between serum retinol concentrations and usual intakes of vitamin A are rare (Hallfrisch et al., 1994; Nierenberg et al., 1997).

Functional impairment has been seen in undernourished children with extremely low serum vitamin A concentrations. For example, results in India (Pirie and Anbunathan, 1981) and Indonesia (Sommer, 1982). showed that at least 75% of the children with xerophthalmia had serum vitamin A concentrations <0.35mol/L. In contrast, in a sample of 252 clinically normal Indonesian children, only 8% had serum vitamin A at this low concentration. Table 18a.1.

Table 18a.1 Serum vitamin A levels (µmol/L) in a sample of Indonesian children with and without ocular lesions. Compiled from Sommer et al., (1980).

		Percent with serum Vit A		
	n	<0.35 µmol/L	0.35–0.69 µmol/L	≥0.7 µmol/L
Normal children	252	8	37	55
Night-blindness or Bitot's spots	325	30	55	15
Corneal xerophthalmia	98	75	24	1

A cutoff for serum vitamin A concentrations <0.70mol/L is a less specific measure of vitamin A deficiency; its ability to predict a deficiency of vitamin A varies widely by region, probably depending on the presence and severity of other risk factors such as infection. Nonetheless, the World Health Organization has guidance on the use of this cutoff (WHO, 2011). For example, in Sri Lanka, serum vitamin A concentrations <0.70mol/L were found in only 28% of children (n=29) with clinical signs of vitamin A deficiency (Bitot's spots, corneal scars, or blindness) and in 5% of the children without positive eye findings (Brink et al., 1979). In Indonesia, 85% of children <6y (n=325) with either night blindness or Bitot's spots had serum vitamin A concentrations <0.70mol/L (Sommer et al., 1980). In contrast, 28% of the children with eye findings and 48% of the children without eye findings had serum vitamin A concentrations >1.05µmol/L. Additional tests may therefore be required to confirm vitamin A depletion at the individual level when a cutoff of <0.70mol/L is used. At a population level, however, a cutoff of <0.70mol/L can be used to indicate whether vitamin A deficiency is likely to be a public health problem (WHO, 2011).

Factors affecting serum retinol concentrations

Age, sex, and race influence concentrations, as indicated by the NHANES II and NHANES III results (Pilch 1987; Stephensen and Gildengorin, 2000; Ballew et al., 2001). African American children and adolescents had lower serum retinol concentrations than Caucasians (Ballew et al., 2001; Neuhouser et al. 2001). In the U.K. National Diet and Nutrition Survey of young people, mean values increased with increasing age in both boys and girls Gregory et al., 2000). In the U.K. Survey of British Adults, mean values were highest for males and females aged 50-64y (Gregory et al., 1990). Age-specific criteria should always be used when interpreting serum retinol concentrations.

Low fat diets (i.e., <5 to 10g daily) impair the absorption of provitamin A carotenoids (Jayarajan et al., 1980), and over the longer-term, lower plasma retinol concentrations. In contrast, the ingestion of meals containing relatively large amounts of either dietary or supplemental vitamin A during the previous 4h does not alter serum retinol concentrations in children or adults (Mejia and Arroyave, 1983; Mejia et al., 1984). Consequently, the collection of fasting blood samples is not necessary for serum retinol determinations, unless retinyl esters and carotenoids are also of interest, in which case fasting is a requirement for accurate interpretation (Mondloch et al., 2o15; Williams et al., 2021).

Other nutrient deficiencies can affect serum retinol concentrations. Protein- energy malnutrition decreases liver apo-retinol-binding protein (RBP) production because of a limited supply of protein substrate (Russell et al., 1983). Consequently, hepatic release of vitamin A is impaired, resulting in decreased serum retinol levels, despite the presence of adequate vitamin A stores in the liver. Zinc deficiency decreases serum retinol concentrations because of its role in the hepatic synthesis or secretion of RBP even in the presence of adequate liver vitamin A stores (Christian and West, 1998) Deficient niacin status sequesters vitamin A in the liver and lowers serum retinol concentrations (Titcomb et al., 2023).

Disease may significantly alter serum retinol concentrations. Chronic renal disease increases serum retinol concentrations by reducing catabolism of vitamin A and its carriers. In contrast, liver disease decreases serum retinol concentrations, probably as a result of a combination of decreased synthesis and secretion of RBP. Cystic fibrosis is also associated with decreased levels of circulating retinol and RBP, probably because of a defect in the transport of vitamin A from the hepatic stores to the periphery.

Infections including HIV (Kafwembe et al., 2001), measles (Thurnham et al., 1997), and parasitic infections (e.g., malaria and *Ascaris lumbricoides*) are associated with low serum retinol concentrations, in some cases because of malabsorption of vitamin A (Sivakumar and Reddy, 1972; 1975). Infections may also cause low serum retinol as a result of transient decreases in the concentrations of acute phase proteins, RBP, and transthyretin, even in the presence of adequate liver vitamin A stores. In NHANES

III, inflammation, indicated by a serum C-reactive protein concentration above 10mg/L, was associated with lower serum retinol concentrations. This, in turn, led to misclassification of vitamin A status (Stephensen and Gildengorin, 2000. The confounding effect of the acute-phase response on serum retinol was not evaluated in earlier NHANES investigations. Clearly, it is important to include a measure of this response to interpret serum retinol concentrations correctly. A variety of methods have been developed to adjust serum retinol concentrations for inflammation using acute phase protein concentrations (Larson et al., 2018). For a practical guide on applying the "Biomarkers Reflecting Inflammation and Nutritional Determinants of Anemia" (BRINDA), see Luo et al., (2023)

Estrogens, either endogenous or those used in contraceptive agents or hormone replacement therapy, increase serum retinol and RBP apparently as a result of increased mobilization of vitamin A from the liver.

Interpretive criteria

Data for mean and selected percentiles for serum retinol concentrations classified by age, sex, and race or ethnicity for NHANES III are presented in Ballew et al. (2001) and IOM (2001). Serum retinol concentrations were also determined in the U.K. national surveys (Gregory et al., 1990; 1995; 2000; Finch et al., 1998). Mean, median, and lower and upper 2.5th or 5th percentiles by age and sex are given.

Age-specific cutoff points for serum retinol concentrations developed for NHANESII are shown in Table 18a.2. These same cutoffs were used for the interpretation of the serum vitamin A values in the Hispanic Health and Nutrition Examination Survey (HHANES) and in NHANES III (Ballew et al., 2001}. In the latter survey, the prevalence of serum retinol <0.70µmol/L was very low for all age and sex groups. However, the prevalence of potentially suboptimal concentrations of serum retinol, defined as <1.05mol/L, was high among certain minority groups of children. Racial and ethnic differences were significant, even after controlling for confounding factors; non-Hispanic African American and Mexican American children were more at risk than were non-Hispanic Caucasian children.

Table 18a.2 Cutoff points recommended by the NHANES II committee for interpreting low serum total vitaminA concentrations in three age categories. Vitamin A status refers to serum and tissue levels of the nutrient. Impairment of function may include impaired dark adaptation, night blindness, and ocular lesions. From Pilch (1987).

Vit. A µmol/L	Age 3–11 y	Age 12–17 y	Age 18–74 y
<0.35	Vitamin A status is very likely to improve with increased consumption of vitamin A. Impairment of function is likely.		
<0.70	Vitamin A status is likely to improve with increased consumption of vitamin A	Vitamin A status is likely to improve with increased consumption of vitamin A; some individuals might exhibit impairment of function.	Vitamin A status is likely to improve with increased consumption of vitamin A; impairment of function is likely.
0.70 - 1.05	Vitamin A status of some subjects may improve with increased consumption of vitamin A. Improvement is most likely in those with values 0.70 - 0.95µmol/L.	Vitamin A status may improve with increased consumption of vitamin A. Improvement is more likely in those with values 0.70 - 0.95µmol/L.	Vitamin A status may improve with increased consumption of vitamin A. Some individuals may exhibit impairment of function.

The U.S. serum retinol cutoff points used for NHANES III are based on their relationship with functional indices of vitamin A deficiency, such as impaired dark adaptation, night blindness, ocular lesions, and possibly impaired immune function and on the serum vitamin A distributions from the earlier NHANES investigations (Pilch 1987). It is not known if these cutoffs are appropriate for other countries.

The World Health Organization uses a single cutoff for serum retinol concentrations of <0.70µmol/L as indicative of a low vitamin A status for children 6-71mo(WHO, 2011). The cutoff for infants <6mo is not defined. Currently, this cutoff is applied irrespective of the prevalence of infection in a particular population.

WHO (2011) has also proposed that prevalence levels of serum retinol concentrations <0.70µmol/L be used to indicate the

possible presence of a public health problem. These define whether vitamin A is either a mild (2% to 9% prevalence), moderate (10% to 19% prevalence), or severe (>20% prevalence) public health problem.

Measurement of serum retinol

Special precautions must be taken in the preparation and storage of serum samples for vitamin A. Serum should be centrifuged soon after the blood is drawn, where possible, and hemolysis and exposure to bright light should be avoided. If separation is not possible within a short time, then the blood sample should be stored in an icebox with ice (Mejia and Arroyave, 1983). If serum is to be stored for prolonged periods of time prior to analysis, it should be quickly frozen in temperature resistant tubes, and then flushed with argon or oxygen-free nitrogen. The tubes should be closed tightly with a screw top, and only a small gas space should be left within the tube. Tubes should be stored, together with a reference serum sample, in the dark at temperatures below –70°C (Olson, 1984). Thawing and refreezing should be kept to a minimum because this may disrupt RBP.

Several methods are available for the analysis of total serum vitamin A or retinol. Only high-performance liquid chromatography can distinguish retinol from the retinyl esters. All the other methods measure total serum vitamin A and generally give comparable results, provided that conditions of collection, storage, and analysis are clearly controlled.

A standard reference material (SRM 968 Fat-Soluble Vitamins in Human Serum) is available from the National Institute of Standards and Technology (Gaithersburg, MD) which is certified for both retinol and retinyl palmitate The within-subject daily CV for serum retinol for subjects following their normal dietary pattern was 11.3% (Gallagher et al., 1992).

High-performance liquid chromatography (HPLC) is the method of choice for the separation and analysis of serum retinol because it has the required precision at low concentrations (de Pee and Dary, 2002). It is also specific and easy to use (WHO, 1996). Usually 100-200µL of serum or plasma are required for analysis. Either reverse- (Bieri et al., 1979; Furr et al., 1984), or normal-phase HPLC is used (Bankson et al., 1986). The former technique is preferred for serum and was used in NHANES III (Ballew et al., 2001). All interfering compounds, such as retinyl esters and phytofluene, can be separated from retinol using HPLC. Moreover, mild hemolysis does not affect serum retinol concentrations measured by HPLC (Marinovic et al., 1997).

Fluorometric methods based on the direct measurement of fluorescence of the retinol-binding complex are used for measuring total serum vitamin A. They are highly sensitive and critical timing is not necessary. Nevertheless, other highly fluorescent substances, particularly phytofluene and phytoene, interfere but can be removed by column chromatography. Alternatively, a correction factor can be used (Thompson et al., 1971). Severe but not moderate hemolysis may be a confounding factor which could lead to serum vitamin A concentrations being underestimated (Marinovic et al., 1997).

A portable fluorometer optimized for vitamin A fluorescence that operates from a direct current power supply (battery, automobile cigarette lighter, or AC/DC power supply) has been developed for field use. It weighs less than 6kg, has no moving parts and uses a cuvet-shaped fluorescent calibration block. Use of this instrument eliminates the necessity for the transport and storage of samples and can generate immediate results provided a portable centrifuge is also available. It reduces the cost of serum retinol analyses (Craft, 2001).

Dried blood spots (DBS) can be used for the measurement of retinol (Erhardt et al., 2002). Positive correlations have been recorded between retinol concentrations in dried blood spots and conventional plasma samples (Figure 18a.3, Craft et al., 2000).

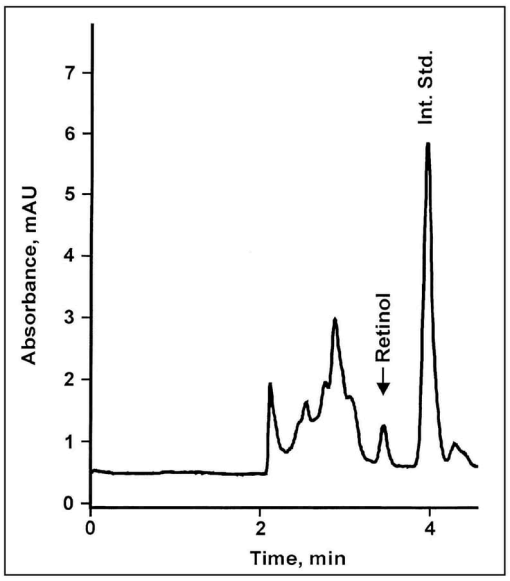


Figure 18a.3 HPLC chromatogram of the separation of retinol extracted from a DBS. (Craft et al., 2000). Retinol and the internal standard, retinyl acetate, elute at 3.6 and 4.0min, respectively. HPLC conditions: Betasil C8, 3µm, 4.6 × 150mm column, methanol / water (95/5 v/v) 1.0mL/min flow rate, 325nm detection.

Blood should be collected from a fingerprick onto filter paper (Schleicher and Schuell 903 specimen collection paper), dried for 3h in the dark, and stored with a desiccant in the dark in bags and at room temperature (i.e., up to 25°C) before analysis. Storage under humid conditions causes loss of retinol in DBS and should be avoided.

Extraction of the DBS retinol can be followed by analysis using normal-phase HPLC with a highly sensitive detector. To reduce the variability of the method, it is recommended that the individual volume of the spots be considered by measuring their sodium content by flame photometry or by weighing the blood spots before extraction. The method has been used on blood spot samples from India, Nepal, Nicaragua, Indonesia, and Guatemala.

18a.2.2 Serum retinol-binding protein

Retinol-binding protein (RBP) is a specific vitamin A transport protein. It is called holo-RBP when it is bound to retinol; the portion without retinol is called apo-RBP. If the liver becomes depleted in retinol, as occurs in the late stages of vitamin A deficiency, RBP accumulates in the liver as apo-RBP, and levels of both serum retinol and RBP decline.

Retinol-binding protein has a single high affinity binding site for a molecule of retinol. The resultant complex of RBP ⋄ retinol, together with one molecule of plasma transthyretin, form a trimolecular complex. Thus, the molar ratio of retinol to RBP in circulation is about one to one. Concentrations of serum RBP may thus be used as a surrogate measure for serum retinol. The assay of RBP is particularly useful in populations where resources and technical support are limited: sample collection and the analytical procedures are easier and cheaper than for serum retinol, and the analysis can be performed on serum from a fingerprick blood sample.

Several studies have confirmed the positive correlation of serum retinol and serum RBP(Solomons et al., 1990; Almekinder et al., 2000; Hix et al., 2004). In a study in the Republic of the Marshall Islands of children who were deficient in vitamin A, Gamble et al. (2001) showed significant correlations between serum retinol and serum RBP (r=0.94) across all retinol concentrations. Severe vitamin A deficiency (serum retinol <0.35µmol/L) was predicted with 96% sensitivity and 91% specificity using serum RBP at 0.48µmol/L as the cutoff, and more moderate vitamin A deficiency (serum retinol <0.70µmol/L) with 87% sensitivity and 98% specificity using a serum RBP cutoff of 0.70µmol/L. Hence, the measurement of serum RBP concentrations appeared to be a practical alternative to using serum retinol in this population. Similar findings were noted for pregnant women in Malawi ((Almekinder et al., 2000).

Nevertheless, a variety of factors may influence the binding of RBP to retinol. These include the presence and magnitude of the acute-phase response, protein-energy malnutrition, liver disease, and chronic renal failure. To overcome these confounding

effects, use of the ratio of serum RBP: transthyretin (TTR) has been proposed (Rosales and Ross, 1998; Rosales et al., 2002). Transthyretin is unaffected by vitamin A status but, like RBP and serum retinol, decreases in infection and trauma. Hence, theoretically, the ratio of RBP:TTR should decline during vitamin A deficiency but not during an acute-phase response, thus enabling these two conditions to be distinguished. Indeed, Rosales et al. (2002) suggest that a cutoff value of 0.36 for RBP:TTR is indicative of marginal vitamin A deficiency. However, use of this ratio has been questioned (Filteau et al., 2000). and more research on the use of the RBP:TTR ratio is warranted.

Interpretive criteria for RBP

There is no consensus on a cutoff value for RBP equivalent to a serum retinol concentration of <0.70µmol/L (de Pee and Dary, 2002). More research on the relationship of serum RBP and serum retinol concentrations in populations with a range of serum retinol concentrations is required before such a cutoff can be defined and is likely dependent on the kit or method used to evaluate RBP. In the interim, investigators are advised to first establish the relationship between serum retinol and serum RBP in a subsample of the population under study by HPLC before using serum RBP as a surrogate for serum retinol concentrations. This is especially important during pregnancy, when both the transport and the metabolism of RBP are altered (Sapin et al., 2000).

Measurement of RBP

Retinol-binding protein, unlike serum retinol, is not photosensitive and is less temperature sensitive and more stable during refrigeration — all attributes that facilitate its use in field surveys.

The RBP assay can be performed using a specific and sensitive radioimmunoassay procedure in which the RBP is bound to radioactively labeled antibodies. This method has a lower limit of detection of 0.5nmol/L (Blaner, 1990; Rosales, 1998). Alternatively, a rapid quantitative enzyme immunoassay (EIA) can be used that has an average intra- and inter-assay variability of 6.7% and 8.0%, respectively (Hix et al., 2004). Tests indicate that the RBP EIA correlates well with serum retinol measured by HPLC. Positive correlations between serum RBP measured by radial immunodiffusion and EIA have been reported (Hix et al., 2004).

An EIA method can also measure RBP in dried blood spots. The kit contains three calibrants over the deficient to normal range for RBP. In the future, it may become feasible to measure RBP directly in a drop of whole blood using a portable fluorometer in the field (Craft, 2001).

A sandwich ELISA method has been developed where RBP, C-reactive protein and α-1-acid glycoprotein can be analyzed in a single run (Erhardt et al., 2004). When properly standardized, this assay could replace serum retinol analysis by HPLC at a fraction of the cost.

18a.2.3 Serum retinyl ester

In normal healthy people, retinyl esters constitute less than 5% of the total vitamin A content in fasting serum samples. However, when the capacity of the liver to store vitamin A is exceeded — for example, after the chronic ingestion of excessive amounts of vitamin A (hypervitaminosis A), or in liver disease — vitamin A is released into the circulation as retinyl esters, and then elevated concentrations of serum retinyl esters are observed. For example, in three patients with hypervitaminosis A, retinyl esters made up 67%, 65%, and 33% of the total vitamin A present in the plasma (Smith and Goodman, 1976).

A cutoff point of retinyl esters >10% of total vitamin A was used to reflect potential hypervitaminosis. Ballew et al. (2001). studied the validity of this cutoff by examining the distribution of concentrations of serum retinyl esters in samples from NHANES III. They noted that 37% of the sample of adults aged 18y had serum retinyl ester concentrations above the cutoff point, but they were unable to find any associations between serum retinyl ester concentrations and five biochemical indices of liver dysfunction. Nonetheless, elevated retinyl esters >7.5% of total was associated with a liver reserve of 3µmol/g (Olsen et al., 2018). and 5% of total was suggested for use in children where more sensitive biomarkers of liver reserves of vitamin A noted hypervitamionsis A (Mondloch et al., 2015; Williams et al., 2021).

Fasting blood samples are essential for serum retinyl ester measurements because concentrations rise transiently after the ingestion of a vitamin A-rich meal or of vitamin A supplements. Measurement of retinyl ester concentrations in serum can be performed using HPLC. The low levels in fasting serum can be measured concurrently with serum retinol concentrations.

18a.2.4 Serum carotenoids

Approximately 50 carotenoids show provitamin A activity and provide about 50% of the total vitamin A intake in the United States, as well as larger percentages in Asia, Africa, and parts of South America. In those countries where dietary carotenoids from plants are the major source of vitamin A and where dietary patterns are relatively constant, serum carotenoids may serve as a useful secondary index of vitamin A intake. For populations receiving most of their vitamin A from animal sources, serum carotenoid concentrations provide no information on vitamin A status.

Major components of serum carotenoids are lutein, zeaxanthin, β-cryptoxanthin, lycopene, α-carotene and β-carotene. Serum carotenoids are of increasing interest because of their possible antioxidant properties and their relationship to the risk of certain cancers, cardioЩvascular disease, macular degeneration, and the onset of cataracts (Omenn et al., 1996).

Unfortunately, several non-nutritional factors are known to influence serum carotenoid concentrations, including age, sex, alcohol intake, physiological state, body mass index, and season (Brady et al., 1996; Neuhouser et al. 2001). Smoking may also modify the relationship of dietary-carotene and serum-carotene levels (Järvinen et al., 1993).

Serum carotenoids were measured following NHANES III. Ford (2000) and Ford et al. (2002) measured serum concentrations of α-carotene, β-cryptoxanthin, lutein, zeaxanthin, and lycopene in samples from NHANES III and presented data for the mean and selected percentiles by age, sex, and race or ethnicity. Serum carotenoid concentrations were not uniformly distributed among the population groups. Children, adolescents, and adults of African descent had the highest total carotenoid concentrations whereas overweight children, Caucasian adolescents, and adults had the lowest.

Plasma β-carotene, α-cryptoxanthin, lycopene, and lutein results are available for the U.K. national surveys (Gregory et al., 1990; 1995; 2000; Finch et al., 1998). Mean, median, and lower and upper 2.5th or 5th percentiles by age and sex are presented.

carotenoids can be both separated and measured by HPLC combined with ultraviolet / visual detection (Sowell et al., 1994) or photodiode array(Mondloch et al., 2o15; Williams et al., 2021). Care must be taken to store frozen serum samples for carotenoid analysis at –70°C or –80°C to avoid decay. A standard reference material for β-carotene can be obtained from NIST (Gaithersburg, MD). The within subject daily CV for serum β-carotene for subjects following their normal dietary pattern is 7.2% (Gallagher et al., 1992).

18a.2.5 Breast milk retinol

Breast milk retinol concentrations can indicate when the maternal vitamin A intake is suboptimal, because lactating women will secrete breast milk with a reduced content of retinol. This usually reflects inadequacies in both maternal dietary intake and body stores of vitamin A. Breast milk retinol concentrations can also be used as an indirect indicator of the vitamin A status of breastfed infants (Stoltzfus and Underwood, 1995).

Most of the vitamin A in breast milk is in the form of retinyl palmitate in the milk fat. Concentrations are very high in colostrum and transitional breast milk (days 7-21 postpartum). After about day 21, the concentration stabilizes, so that breast milk samples taken after the first month postpartum are the most useful. The concentration of retinol in breast milk varies over the course of a feed: the lowest concentration is in the first milk expressed from a full breast because the fat content is highest at the end of a feed. Hence, preferably all the milk from a full breast that has not been used to feed an infant for at least 2h should be collected.

Stoltzfus et al. (1993). used full breast milk samples to investigate the effect of high-dose vitamin A supplementation on the vitamin A status of mothers and their infants in Indonesia. At 1-8mo postpartum, the breast milk retinol concentrations of the supplemented mothers were significantly higher than in the placebo group (Figure 18a.4).

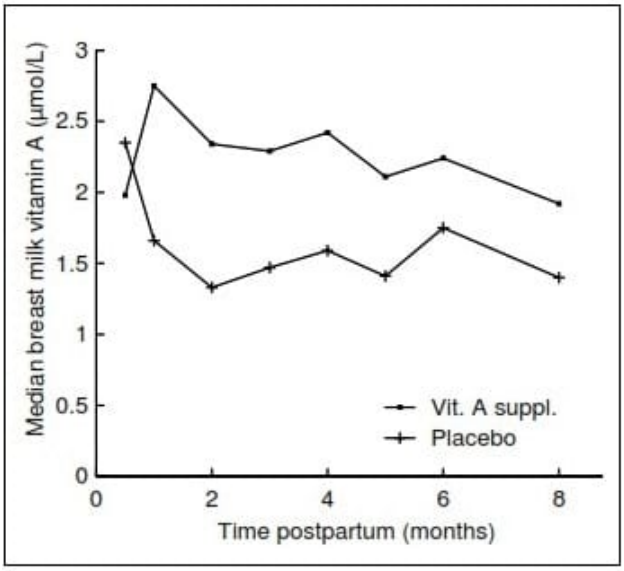


Figure 18a.4 Retinol concentrations (µmol/L) in breast milk at baseline and during the subsequent 8mo in supplement and placebo groups. Data from (Stoltzfus et al., 1993).

In practice, collecting the entire contents of one breast is difficult to achieve. Instead, a standardized collection procedure can be used to assess vitamin A status at the individual level. This involves collecting a small sample of breast milk (8-10mL) from a full breast before the infant starts suckling, either by manual self-expression or by using a breast pump (Stoltzfus and Underwood, 1995).

For population assessment, casual breast milk samples can be taken randomly throughout the day and at varying times after the infants were last fed. This will ensure that the variation in milk fat is randomly sampled. However, if random sampling is not possible, then the breast milk retinol concentrations can be expressed in terms of the fat content (Table 18a.3). (WHO, 1996). In a study of rural women from Bangladesh, Rice et al. (2000) noted that breast milk vitamin A concentrations expressed per gram of milk fat were a particularly responsive indicator of vitamin A status.

Table 18a.3 Response to postpartum vitamin A supplementation measured by maternal and infant indicators. The values shown are means ±SD. A natural log transformation was used in all cases to improve normality except for the serum retinol data. The means and SDs of the transformed values are presented. [n], number of samples. Data from Rice et al., 2000.

Indicator (month post-partum)	Vitamin A group [n]	Placebo group [n]	Standardized difference
Breast milk vit.A (µg/g fat) in casual samples (3 mo)	2.05 ± 0.44 [36]	1.70 ± 0.47 [37]	0.76
Breast milk vit.A (µmol/L) in casual samples (3 mo)	0.12 ± 0.70 [36]	−0.18 ± 0.48 [37]	0.50
Maternal serum retinol (µmol/L) (3 mo)	1.45 ± 0.47 [34]	1.33 ± 0.42 [35]	0.27
Breast milk vit.A (µmol/L) in full samples (3 mo)	−0.33 ± 0.74 [33]	−0.45 ± 0.53 [35]	0.19
Breast milk vit.A (µg/g fat) in full samples (3 mo)	1.87 ± 0.51 [33]	1.82 ± 0.45 [35]	0.10

There are two circumstances when it may be preferable to express breast milk retinol concentrations per unit volume rather than per gram of fat. First, when a field survey includes mothers with a wide range in the stage of lactation, there will be a large standard deviation around the mean retinol concentrations because of the marked variation in milk retinol concentration from early to late lactation. This large variation may be masked by expressing the retinol concentration in terms of fat content. Secondly, in studies

in which both the β-carotene and the fat content of breast milk are increased because of a vitamin A intervention, retinol concentrations expressed per unit volume will provide a better measure of the response than when expressed in terms of the fat content (de Pee et al., 1997).

Interpretive criteria

The average retinol content of breast milk from vitamin A-sufficient lactating women ranges from 1.7 to 2.5μmol/L; values are often <1.4μmol/L for vitamin A-deficient mothers. The cutoff value defined by WHO (1996) at which both mother and breastfed infant are likely to be at risk for vitamin A deficiency is <1.05μmol retinol/L or <28nol/g milk fat. The sensitivity of breast milk retinol using <1.0μmol/L milk was 42% using total liver reserves of 0.10μmol retinol/g as the reference standard for definition of vitamin A deficiency in Zambian women (Kaliwile et al., 2021).

WHO (1996) defined two population prevalence levels that are indicative of a public health problem for breast milk concentrations. When 10% to <25% of the sample population of mothers have breast milk retinol concentrations <1.05 μmol/L, then the public health problem for vitamin A deficiency is classified as moderate; when the prevalence is >25%, the problem is severe.

Measurement of breast milk retinol

After collecting breast milk samples, they must be placed immediately in an insulated ice box for transport from the field to the base laboratory. If the analysis cannot be completed immediately, samples may be frozen below –20°C. Care must be taken to ensure that the thawed samples are thoroughly mixed before aliquots are removed for analysis into amber or yellow polypropylene tubes with airtight caps. Analysis of retinol in breast milk can be performed by HPLC, after saponifying the breast milk samples. The fat content of breast milk can be determined in the field by using the creamatocrit method. This involves collecting a sample (about 75μL) of well-mixed breast milk into a hematocrit capillary tube, which is then sealed at one end. The tubes are spun in a hematocrit centrifuge, after which the length of the cream layer is measured. The amount of fat is determined by comparison with a standard curve of fat concentrations compiled using a standard lipid assay (Ferris and Jensen, 1984).

18a.2.6 Relative dose response

The relative dose response (RDR) test can be used to estimate the liver stores of vitamin A and thereby identify individuals with marginal vitamin A deficiency. The test is based on the observation that during vitamin A deficiency, when liver stores are diminished, RBP accumulates in the liver as apo-RBP. Following the administration of a test dose of vitamin A, some of the vitamin A binds to the excess apo-RBP in the liver. It is then released as holo-RBP (i.e., RBP bound to retinol) into the circulation (Loerch et al., 1979). Consequently, in vitamin A- depleted individuals, there is a rapid and sustained increase in serum retinol after a small oral dose of vitamin A, whereas in individuals with normal liver vitamin A stores, this rise in serum retinol is either very small or does not occur.

The validity of the RDR test as an index of body stores of vitamin A has been studied by comparing vitamin A concentrations in liver biopsy samples with corresponding RDR test results for otherwise healthy surgical patients (Amédée-Manesme et al., 1984; 1987). Of the twelve surgical patients, the two with the lowest liver vitamin A concentrations had the highest RDR values. Following supplementation with vitamin A, RDR values fell to <5%. Patients with liver vitamin A concentrations ranging from 0.2 to 1.5μmol/g had RDR values from 0% to 12% (Amédée-Manesme et al., 1984).

In a study of Brazilian children from low-income families, all the children with serum retinol concentrations <0.70μmol/L had elevated RDR values. Moreover, 86% of the children with serum retinol concentrations of 0.74-1.02μmol/L, and 26% with serum retinol concentrations of 1.05-1.40μmol/L also had elevated RDR values Flores et al., 1984) (Table 18a.4).

Table 18a.4 Percentage of children with a positive RDR test classified by serum retinol levels. From Flores et al. (1984)

Serum retinol (μmol/L)	%	No. tested
≤0.70	100	12
0.71 - 1.04	86	21

Serum retinol (µmol/L)	%	No. tested
1.05 - 1.39	26	19
>1.40	3	39
Total tested		91

After supplementation with vitamin A, all the elevated RDR values reverted to normal. These results indicate that the RDR is a more sensitive index of marginal vitamin A status than using serum vitamin A concentrations <0.70µmol/L, but it does not distinguish among different levels of adequate vitamin A reserves (Solomons et al., 1990).

Other factors associated with low RDR values include malabsorption, infection, liver disease, severe protein-energy malnutrition, and zinc deficiency. Such factors reduce the sensitivity and the specificity of the RDR test (Mobarhan et al., 1981; Russell et al., 1983. Campos et al., 1987). For example, when an oral dose of vitamin A was given to patients with varying degrees of liver dysfunction and protein-energy malnutrition, no correlation was observed between the vitamin A content of liver biopsies and the RDR test result (Mobarhan et al., 1981; Russell et al., 1983). These results were attributed to malabsorption of the oral dose because when an intravenous injection of retinyl palmitate was given to children with liver disease, the RDR test proved to be a reliable and sensitive index of vitamin A status (Amédée-Manesme et al., 1987).

Protein deficiency interferes with the RDR test by decreasing liver synthesis of the rapid turnover apo-RBP, while zinc deficiency interferes with the liver synthesis or secretion of RBP (Christian and West, 1998). A further limitation of the RDR test is low precision, which may lead to serious errors in classifying subjects with vitamin A deficiency, especially in populations where the prevalence is moderate (Solomons et al., 1990).

Interpretive criteria for the RDR test

Vitamin A-replete subjects have RDR values ranging from zero to 14%. Relative dose response values >20% are generally considered indicative of inadequate hepatic stores of vitamin A and marginal vitamin A status (WHO, 1996). When >20% to <30% of the sample population show abnormal RDR values (i.e., >20%), then a public health problem of moderate importance may be assumed. When the prevalence is >30%, the public health problem is severe (WHO, 1996). A systematic review and meta-analysis revealed 68% sensitivity and 85% specificity for the RDR test using total liver reserves of 0.10µmol retinol/g as the reference standard for definition of vitamin A deficiency (Figure 18a.5) (Sheftel and Tanumihardjo, 2021).

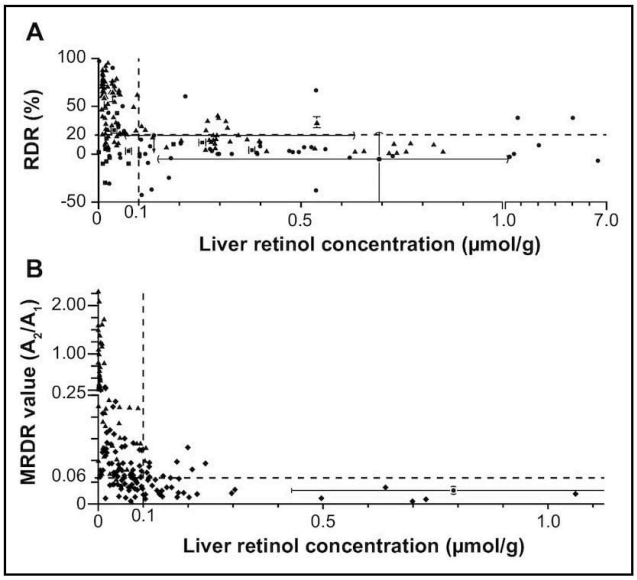


Figure 18a.5 The dose response tests compared with liver retinol concentrations from multiple studies. The symbols represent individual or group values, triangles: rats, squares: calves, diamonds: swine, and circles: humans. A relative dose response (RDR) value of >20% and an A2/A1 for the modified relative dose response (MRDR) test of >0.060 are considered vitamin A deficiency

Measurement of relative dose response

Two blood samples are needed for the RDR test. A baseline fasting blood sample is taken immediately before the administration of a small oral dose (450 to 1000µg) of vitamin A (as retinyl acetate or retinol palmitate). Next, a high-fat snack that contains minimal vitamin A is consumed to ensure absorption of vitamin A. The second blood sample is taken 5h later. The RDR (%) is calculated as:

$$\frac{\text{plas. retinol (5 h)} - \text{plas. retinol (0 h)} \times 100\%}{\text{plas. retinol (5 h)}}$$

Concentrations of serum retinol for the RDR test should be measured by reverse-phase HPLC because this method has the required precision at low concentrations. Even a small analytical error can alter the RDR value significantly, especially when serum retinol concentrations are low.

18a.2.7 Modified relative dose response

The modified relative dose response (MRDR) test requires only one blood sample, avoiding the necessity of taking a baseline sample. This modified test has been used to assess the vitamin A status in children and in pregnant and lactating women (Tanumihardjo et al., 1990; 1994; 1995; 1996).

For the test, a small oral dose of vitamin A is administered, usually 3,4-didehydroretinyl acetate (DRA). This is followed by a high-fat, low-vitamin A snack to facilitate the absorption of the DRA. The latter combines with the RBP in the same way as retinol but, unlike retinol, is not normally found in human plasma except when high levels of freshwater fish are consumed. The DRA is hydrolyzed in the gastrointestinal tract to 3,4-didehydroretinol (DR), absorbed, and reesterified, mainly with palmitic acid in the intestinal mucosal cells (Tanumihardjo et al., 1995). The serum concentrations of DR and retinol (R) in the single blood sample are measured by HPLC. The molar ratio of [DR]:[R] in the blood sample is a measure of the response.

Interpretive criteria for the MRDR

Ratios >0.060 are considered indicative of marginal vitamin A status (subclinical deficiency), whereas those <0.030 are satisfactory (Tanumihardjo et al., 1996). The WHO (1996). criteria for a moderate public health problem are a prevalence of MRDR ratios >0.060 of between 20% and 30%. If the prevalence is >30%, the public health problem is severe. Because of the preferred use of the MRDR in subsets of population studies, these criteria are being reexamined.

A limitation is that DRA is not yet commercially available and presently must be synthesized in the laboratory. Nonetheless, the test has been used in several population health surveys (Faye et al., 2021; Williams et al., 2021; Suri et al., 2021). A systematic review and meta-analysis revealed 80% sensitivity and 69% specificity for the MRDR test using total liver reserves of 0.10µmol retinol/g as the reference standard for definition of vitamin A deficiency (Figure 18a.5) (Sheftel and Tanumihardjo, 2021).

Measurement of the MRDR

The recommended doses of DRA for use in the field are 3.0µmol for infants, 5.3µmol for preschool children (2 to 6y), 7.0µmol for preadolescent children (6-12y), and 8.8µmol for teenagers and adults (>12y) (Tanumihardjo et al., 1996). The acceptable time between administering the oral DRA dose and obtaining a single blood sample (about 0.5-2ml) is 4-7h. To enhance the stability of the DRA dose, it should be dissolved in corn oil and stored in amber vials inside a cooler on ice in the field, or at –20°C to –70°C for long-term storage. The blood sample should also be stored on ice in a light-protected cooler after collection, prior to separation of the serum. Analyses of DR and retinol (R) concentrations in the serum should be performed by HPLC. The molar ratio of [DR]:[R] (sometimes reported as A2/A1) is then calculated.

18a.2.8 Subjective assessment of night blindness

Night blindness, or the inability to see after dusk or at night, has been reported among young children and women of reproductive age in developing countries with moderate to severe vitamin A deficiency (Escoute et al., 1991; Christian et al., 1998). It is the most common ocular manifestation of vitamin A deficiency and is often described by specific terms in countries or cultures where the prevalence is high. For example, in some cultures it is known as "chicken eyes" or "chicken blindness" (Christian P. 2002.). Poor

dark adaptation resulting in night blindness arises when there is reduced production in the rods of the visual pigment rhodopsin, or opsin protein bound to the retinal form of vitamin A.

To assess the prevalence of night blindness, the local term for night blindness must first be identified through focus group discussion (Dawson et al., 1993), and its reliability field tested. Next, a night blindness history is elicited via interviews. Care should always be taken during this stage to exclude those individuals whose night blindness results from other causes, such as the rare hereditary eye disease retinitis pigmentosa.

Table 18a.5 Mean serum vitamin A level in Indonesian preschool children by clinical indicators of vitamin A status: night blindness and conjuctival xerosis with Bitot's spots. Data from Sommer et al., (1980).

Clinical status	*n*	Mean (µmol/L)
Night blindness reported; no conjunctival xerosis or Bitot's Spots	174	0.49
Controls	161	0.62
No night blindness; conjunctival xerosis and Bitot's Spots present	51	0.47
Controls	45	0.60
Night blindness reported; with conjunctival xerosis and Bitot's Spots	79	0.42
Controls	76	0.64

Some studies of children >2y have been based on interviews to elicit any history of night blindness. In general, reliable data on night blindness cannot be determined for children <2y in this way because they are not able to move around freely after dusk and bump into objects. In studies of children >2y, WHO (1996) recommends the use of specific questions to increase the specificity and reduce misclassification of self-reported night blindness. Sommer et al. (1980) reported that 85% of Indonesian preschool children with reported night blindness had low serum vitamin A concentrations (i.e., <0.70µmol/L). This strong association between night blindness and serum retinol levels is shown in Table 18a.5. This also shows similar positive findings for a second clinical sign of vitamin A deficiency — Bitot's spots with conjunctival xerosis.

Determination of the prevalence of night blindness is recommended as a tool to assess vitamin A deficiency in women of reproductive age. When using maternal night blindness in this way, only women with a previous pregnancy that ended in a live birth in the past 3y should be included (Christian P. 2002). Significant associations between maternal night blindness and low serum and breast milk retinol concentrations, as well as functional indices such as abnormal CIC and impaired dark adaptation (Section 18.1.9), have been reported among pregnant Nepalese women, as shown in Table 18a.6.

Table 18a.6 Association of night blindness with other indicators of vitaminA deficiency in pregnant women in Nepal. N, sample size; n, number of subjects below cutoff; OR, odds ratio; CI, confidence interval; CIC, conjunctival impression cytology. From Christian, (2002).

Vitamin A status indicators	Night blind N n(%)	Not night blind N n(%)	OR	95% CI
Serum retinol <0.7 µmol/L	85 44 (51.0)	90 19 (21.1)	4.0	2.2–7.4
Serum retinol <1.05 µmol/L	85 65 (76.5)	90 50 (55.5)	2.5	1.4–4.6
CIC abnormal	85 24 (28.2)	90 11 (12.2)	2.8	1.3–6.1
Dark adaptation abnormal	94 67 (71.2)	98 42 (43.8)	3.3	1.8–6.0
Breast milk vit. A <1.05 µmol/L	94 56 (59.6)	97 41 (42.3)	2.0	1.1–3.6

In a large population-based control study in Nepal, women with night blindness were also less likely to consume dietary sources of vitamin A (Christian et al., 1998). In a randomized placebo-controlled supplementation trial in Nepal, women in the placebo group with night blindness during pregnancy had a mortality rate four times higher than did women without night blindness. This mortality level was reduced by 68% after weekly vitamin A supplementation (Christian et al., 1998).

Interpretive criteria

A recommended cutoff at which vitamin A deficiency is considered a significant public health problem within a community is when the prevalence of maternal night blindness is 5% or greater (Christian, 2002). This cutoff value takes into account the misclassification that may occur (i.e., false positives) among women who report problems with daytime vision (i.e., daytime blindness). Efforts should be made during surveys to exclude those women whose night blindness is likely due to daytime blindness. WHO (1996) have proposed cutoffs for the prevalence of night blindness for children 24-71mo. Three recommended cutoffs for the prevalence of night blindness to define a public health problem and its level of importance (i.e., mild, moderate, or severe) are given; these are shown in Table 18a.7 together with minimum sample sizes required. However, the large minimum sample sizes required together with the increasing use of vitamin A supplements among young children in most developing countries preclude the use of these WHO cutoffs for children. Instead, in most low-income countries, the assessment of nightblindness in women of reproductive age is probably more useful.

Table 18a.7: prevalence of night blindness in children 24-71mo, and the minimum sample sizes for identifying a mild, moderate, or severe vitamin A deficiency public health problem with a relative precision of 20% and 50% at the 95% confidence level. From WHO (1996).

	prevalence	Minimum sample for 20% relative precision	Minimum sample for 50% relative precision
Mild	<1%	—	—
Moderate	≥1% to <5%	4706	753
Severe	≥5%	1825	292

18a.2.9 Rapid dark adaptation test

Before night blindness develops, disturbances in dark adaptation occur. These can be detected by specially designed noninvasive tests (Congdon et al., 1995). The conventional laboratory-based, formal dark adaptometry test is a tedious and time- consuming procedure. Instead, a rapid dark adaptation test (RDAT), suitable for field conditions, has been developed. This is based on the measurements of the timing of the Purkinje shift (Thornton, 1977), in which the peak wavelength sensitivity of the retina shifts from the red toward the blue end of the visual spectrum during the transition from photopic or cone-mediated day vision to scotopic or rod-mediated night vision. This shift causes the intensity of blue light to appear brighter than that of red light under scotopic lighting conditions.

The RDAT requires a light-proof room, a light source, a dark, nonreflective work surface, a standard X-ray view box, and sets of red, blue, and white discs; details are given in Vinton and Russell (1981).

Measurements for the RDAT are undertaken during the first few minutes of dark adaptation. This is a disadvantage because the measurements rely mainly on the cones or light vision cells in the retina instead of the rods or dark vision cones. As a result, the test is not very sensitive to the early signs of vitamin A deficiency (Favaro et al., 1986; Kemp et al., 1988). An additional disadvantage is that the RDAT is not appropriate for preschool children, who are too young to perform the test accurately. This is unfortunate because preschool children are the group most at risk for vitamin A deficiency. False positives may also occur and between examiner variability may produce inconsistencies in the results.

Age influences dark adaptation and hence must be taken into account when determining the normal range of the rapid test adaptation times for healthy reference populations (Vinton and Russell, 1981).

18a.2.10 Pupillary and visual threshold test

A scotopic (dim light vision) device has been developed to assess the responsiveness and sensitivity of the pupil to light as an indication of an individual's dark adaptation threshold. The test measures the threshold of light at which pupillary contraction occurs under dark-adapted conditions (Congdon et al., 1995). A portable field dark adaptometer was developed that includes a digital camera, retinal bleaching flash, and Ganzfeld light source housed within a pair of light-obscuring goggles (Labrique et al., 2015). Pupillary response measurements may be more affordable for countries wishing to use this device.

Unlike the RDAT, the pupillary and visual threshold test can be conducted in the field on individuals of all ages, including preschool children, who cannot be tested by the RDAT. The test requires minimal cooperation from the subjects and takes about 20 min per subject. A darkened facility is required for this test. A portable tent has been developed and tests devised to ensure that the darkness in the testing area is adequate (Sanchez et al., 1997). The apparatus consists of two handheld illuminators, each having a yellow-green, light-emitting diode light source (dominant wavelength = 572nm) with 12 intensity settings. Each illuminator is designed to fit entirely over one eye and to illuminate the entire retina. One of the illuminators is designed to measure light of "low intensity" (illumination range: 8.75 to -3.00 log cd/m^2) and the other "high intensity" (illumination range -4.16 to 0.44 log cd/m^2).

Before testing, participants are subjected to a camera-flash "partial bleach" of the full retina, which involves placing the subjects in a dark room and exposing them to a camera flash reflected through a foil-lined cone, after which they are allowed 10 min of dark adaptation. The visual threshold is measured first by placing the low-intensity illuminator over the subject's left eye. The light intensity is then incremented over 11 intensity settings (roughly a 4-log unit range) at 10s intervals until a pupillary response (i.e., quick contraction of the pupil on presenting the stimulus) is seen in the uncovered right eye on two successive trials. The uncovered right eye is observed with an obliquely mounted red LED light source (dominant wavelength = 626nm), which preserves dark adaptation in both the subjects and the observer. Next, the pupillary threshold is measured, as described above, using the high intensity illuminator; further details are given in Congdon et al. (1995; 2000). and in Christian et al. (2001). All tests should be performed using standardized procedures and well-trained examiners, as discussed in Christian et al. (2001).

The stimulus for the visual or the pupillary threshold is defined as the lowest level at which the subject can correctly distinguish stimulus from non-stimulus on three successive trials. High pupillary and visual scores reflect a pupillary response achieved at a greater light intensity, and indicate poorer dark adaptability (Congdon et al., 1995).

Additional studies are needed to establish whether testing for both pupillary and visual thresholds is necessary, because the latter requires less training and standardization of personnel. In some studies on young children aged 1 to 2y, only pupillary testing has been performed because complete visual testing on such young children is not always possible (Congdon et al., 1995).

The pupillary and visual threshold tests have been validated as an index of vitamin A status using controlled vitamin A supplementation trials on children and pregnant women (Congdon et al., 1995; 2000; Congdon and West, 2002). Significant improvements in dark adaptation as assessed by visual and/or pupillary testing were reported in those subjects supplemented with vitamin A but not with a placebo. Moreover, dark adaptation scores, measured by visual and/or pupillary testing, correlate well with serum retinol, as shown in Figure 18a.6, and with RDR (Congdon et al., 1995; 2000; Sanchez et al., 1997).

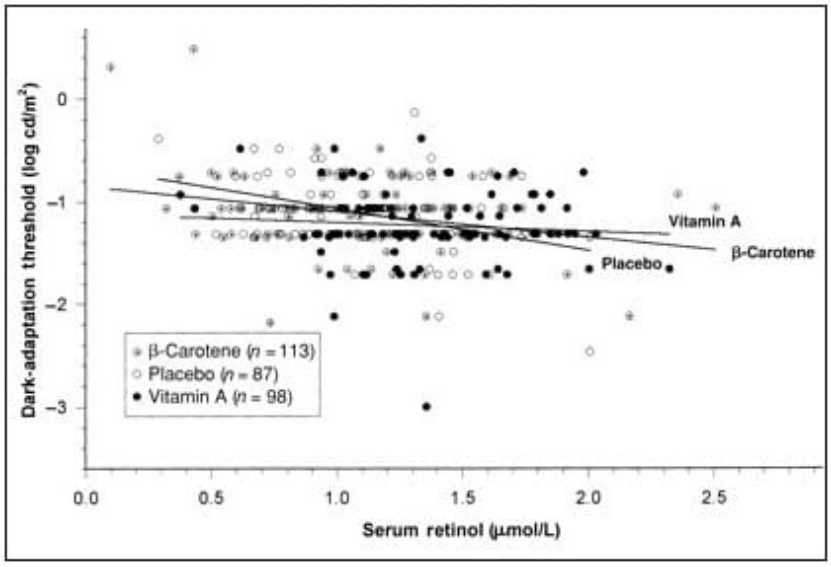


Figure 18a.6 Dark-adaptation threshold and serum retinol concentration of pregnant Nepali women receiving β-carotene, vitamin A, and placebo. Dark-adaptation threshold and serum retinol concentration were inversely correlated in the placebo (regression coefficient = −0.014, P = 0.0005) and β-carotene (regression coefficient = −0.009, P = 0.005) groups. Modified from Congdon et al. (2000).

Nevertheless, there are limitations associated with the pupillary and visual threshold tests. For example, some investigators have shown overlapping values for deficient and nondeficient population groups, so that scotopic threshold testing may be more suitable for population assessment than for individuals. Further, the time required to administer the test is about 20 min per subject, so it is probably not feasible to test large numbers of individuals. Sanchez et al. (1997) suggest that for populations at high risk for vitamin A deficiency, as few as six subjects may be sufficient to show that the group mean thresholds differ significantly from normal. In populations with mild vitamin A deficiency, testing about 100 subjects is probably adequate (Sanchez et al., 1997).

Another limitation is that recovery of the normal pupillary response after dosing with vitamin A supplements takes about 4-6wk. One must wait at least this long before retesting treated individuals. Finally, the period required to train examiners to recognize the pupillary response is 1-3d.

Two cutoff values were proposed to indicate whether vitamin A deficiency is a problem or not in an area, based on a mean pupillary dark adaptation score for a population. Scores worse than the cutoff of –1.11 log cd/m^2 are said to indicate vitamin A deficiency, whereas scores better than the normal cutoff of –1.24 log cd/m^2 should indicate that a population is normal, or that an intervention has successfully improved vitamin A status (Congdon and West, 2002). The population mean should be calculated for this test, rather than the proportion falling below a specific value, as has been stipulated for the other tests of vitamin A nutriture. This approach is recommended because in populations with marginal vitamin A status, the small proportion of subjects with abnormal values for the test could potentially markedly increase the sample size required for testing.

18a.2.12 Stable isotope methods and total body stores of vitamin A

Retinol isotope dilution (RID) tests are the only method that indirectly provide a quantitative measure of the hepatic stores of vitamin A. It involves the administration of an oral dose of isotopically labeled vitamin A. The dose is allowed to equilibrate with the vitamin A pool in the body. A post-dose blood sample is taken, and the ratio of labeled to non-labeled vitamin A in serum is measured by mass spectrometry. The amount of total body stores of vitamin A is related to the extent of dilution of the labeled tracer and is calculated using prediction equations (Gannon and Tanumihardjo, 2015).

The isotope dilution method was validated in adult surgical patients in the United States and Bangladesh by measuring liver vitamin A biopsy samples (Furr et al., 1989; Haskell et al., 1997). For example, in ten U.S. surgical patients, the correlation coefficient between calculated and measured liver vitamin A concentrations was 0.88.

The length of time required for the isotopic dose to equilibrate with the vitamin A pool in the body varies according to the age and

total body vitamin A stores of the study group. Equilibration periods after the test dose ranging from 11 to 26d have generally been used (Furr et al., 1989; Haskell et al., 1998; Ribaya-Mercado et al., 1999). although Tang et al. (2002) suggest that 3d may suffice. A shorter equilibration time facilitates the use of this isotope-dilution method in field settings, but may not offer an accurate quantitative estimate. The MRDR test should be considered if enough time is not available in the field to obtain the post-dose sample.

The RID test has been used mostly in research settings, but the application for use in subsets of population surveys is on the horizon (Tanumihardjo, 2020). The International Atomic Energy Authority has assisted many low-income countries on its use (Sheftel et al., 2018). A combined analysis from five African countries revealed stark differences in vitamin A status based on total liver vitamin A reserves estimated by RID. Ethiopia was experiencing vitamin A deficiency, while South Africa had a high rate of hypervitaminosis A (Suri et al., 2023).

Interpretive criteria

The RID test using appropriate assumptions, gives a prediction of the total liver vitamin A reserves (Gannon and Tanumihardjo, 2015). Therefore, a value that represents vitamin A status is determined, which reflect that actual hepatic reserve. While best used at the group level, vitamin A status can be viewed across the continuum, such as that represented in Figure X.

Hepatic liver reserves to define vitamin A deficiency and toxicity

The criteria to define vitamin A deficiency were evaluated in light of the Dietary Reference Intakes (DRIs) for North Americans (Tanumihardjo, 2021). The criteria used to formulate DRIs were clinical eyes signs, circulating plasma retinol concentrations, bile excretion containing vitamin A metabolites, and long-term vitamin A storage. In consideration of DRI criteria, induced biliary excretion and long-term vitamin A storage do not occur until liver vitamin A concentrations are >0.10µmol/g (Tanumihardjo, 2021). Therefore, in line with recommendations of an expert group working on Biomarkers of Nutrition for Development (Tanumihardjo et al., 2016), vitamin A deficiency should be defined as <0.10 µmol/g liver.

Direct measure of hepatic vitamin A concentrations are only feasible in special circumstances, either during surgery (Furr et al., 1989; Haskell et al., 1997). or on cadavers (Olsen et al., 2018). In children (n=366) dying from multiple causes, vitamin A deficiency was found in 34.2% and hypervitaminosis A was determined in 8.7% (Gupta et al., 2024). This underlines the fact that vitamin A deficiency is still a public health concern in young children.

18a.2.13 Multiple indices

Vitamin A deficiency disorders are not defined with any certainty using a single measure of vitamin A status. WHO (2011). has recommended a combination of biochemical, functional, and clinical indicators for children aged 6-71mo; the indicators are given in Table 18a.8. together with the corresponding cutoff values used in population assessment. Some caution is needed when applying these cutoffs to individuals because of the influence of the many confounding factors. Table 18a.8 also provides information on the population prevalence levels of each indicator for defining a mild, moderate, or severe public health problem in relation to vitamin A.

Table 18a.8: Indicators of subclinical vitamin A deficiency in children 6–71mo of age and their use in detecting a public health problem. The prevalence in a population of at least two of the above indicators of vitamin A status, at levels below the cutoff, indicates a public health problem. RDR, relative dose response; MRDR, modified relative dose response. From (WHO, 1996).

Indicator (cutoff)	Mild prevalence below cutoff	Moderate prevalence below cutoff	Severe prevalence below cutoff
Night blindness present	>0 to <1%	≥1% to 5%	≥5%
Serum retinol ≤ 0.70 µmol/L	≥2% to <10%	≥10% to <20%	≥20%
Breast milk retinol ≤ 1.05 µmol/L or ≤ 0.028 µmol/g (≤ 8µg/g) milk fat	<10%	≥10% to <25%	≥25%
RDR ≥20%	<20%	≥20% to <30%	≥30%
MRDR ratio ≥0.06	<20%	≥20% to <30%	≥30%

According to WHO (2011). a public health problem exists when a population has at least two biological indicators with a prevalence above the level corresponding to a deficiency, or one biological indicator of deficiency supported by a composite of at least

four demographic and ecological risk factors, two of which must be nutrition- and diet-related. Details of these demographic and ecological risk factors are given in WHO (2011).

The selection of the most appropriate combination of criteria depends on the purpose of the study, the expected range of vitamin A status for the study group, and the resources available. WHO often uses mortality outcomes to define its guidance, especially for young children (WHO, 2011).

This page titled 15.2: Indices of vitamin A status (18a.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16: VITAMIN D (CHAPTER 18B)

Abstract

Vitamin D is a generic term for a group of related substances, including vitamin D_2 (ergocalciferol) and vitamin D_3 (cholecalciferol). It is a precursor to the steroid hormone 1,25-dihydroxyvitamin D (calcitriol) that helps the body absorb and retain calcium and phosphorus, thus contributing to bone health. Deficiencies in vitamin D may cause rickets, osteomalacia and osteoporosis. Sun-mediated synthesis of vitamin D in the skin may be important in lower latitudes, but in settings where this synthesis is not significant, vitamin D is an essential nutrient. As vitamin D is a fat-soluble vitamin, deficiencies may be caused by malabsorption syndromes. Exposure and status of vitamin D is assessed by measuring the concentration of the transport metabolite, 25-hydroxyvitamin D, that reflects both ingested and skin-synthesized sources. 1,25-dihydroxyvitamin D (calcitriol) is not a measure of vitamin D status as levels in blood reflect temporary stimulation and not nutrient stores. Other measures of vitamin D status include factors and molecules related to its role in bone development; little is known about specific biomarkers of the nonskeletal roles of vitamin D

Vitamin D (calciferol) is a generic term for a group of related fat-soluble substances, including vitamin D_2(ergocalciferol), vitamin D_3 (cholecalciferol) and their metabolites (Figure 18b.1).

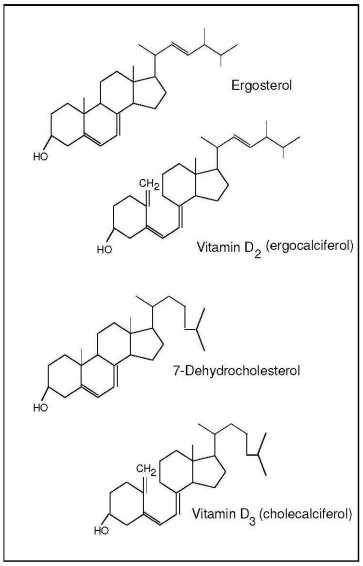


Figure 18b.1 The structures of two forms of vitamin D (Vitamin D_2 and Vitamin D_3) and their precursor sterols.

Vitamin D_2 is derived from the plant sterol, ergosterol, and is the form used to fortify vegan foods and in some high-dosage pharmaceutical preparations. Vitamin D_3originates from the action of ultraviolet B (UVB) rays from sunlight or artificial sources on its precursor sterol, 7-dehydrocholesterol. This is present in the skin and some animal-based dietary sources, including some fish liver oils, the flesh of fatty fish, eggs, and organ meats. Foods may be fortified with vitamin D (e.g., milk, cereals, bread, margarine) in some countries (Cashman and O'Dea, 2019).

16.1: Functions of vitamin D (18b.1)

16.2: Metabolism of vitamin D (18b.2)

16.3: Vitamin D deficiency in humans (18b.3)

16.4: Food sources and dietary intakes (18b.4)
16.5: Nutrient reference values for vitamin D (18b.5)
16.6: Biomarkers of vitamin D status (18b.6)
16.7: Interpretive criteria (18b.8)
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)
16.10: Other vitamin D metabolites (18b.11)
16.11: Serum parathyroid hormone (18b.13)
16.12: Calcium and phosphorus in serum and urine (18b.14)

This page titled 16: Vitamin D (Chapter 18b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.1: Functions of vitamin D (18b.1)

The active form of vitamin D, 1,25-dihydroxyvitamin D (calcitriol), primarily functions as a steroid hormone, ensuring adequate intestinal absorption of calcium and phosphorus and regulating bone mineralization. In this role, the pathway of synthesis of the active form 1,25-dihydroxyvitamin D (calcitriol), is endocrine as it is made in a tissue only when it is required, then circulates to target tissues, where it is present for only a short period of time.

Vitamin D also has a critical role in a number of cellular functions in many tissues in the body, including bone, placenta, prostate, keratinocytes, macrophages, T-lymphocytes, dendritic cells, parathyroid gland and some cancers (Wacker and Holick, 2013). The active metabolite 1,25-dihydroxyvitamin D (calcitriol) is involved in modulating the immune response and in regulating cell differentiation, proliferation, and apoptosis (Norman, 2008). In these non-skeletal roles, the pathway of synthesis of this active form is autocrine / paracrine, whereby the active form stays in the cell or is transported locally.

This page titled 16.1: Functions of vitamin D (18b.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.2: Metabolism of vitamin D (18b.2)

Vitamin D can be obtained in two ways: from skin synthesis of vitamin D3 (cholecalciferol) and from ingestion of the parent compounds (D2 (ergocalciferol) or D3) from foods or supplements. This situation creates problems in assessing vitamin D status as dietary intake alone is not sufficient to gauge risk for deficiency. Sun exposure along with behavioural and environmental factors affecting skin synthesis must also be taken into account.

18b.2.1 Skin synthesis of vitamin D_3 (cholecalciferol)

The requirement for vitamin D can be met by skin synthesis alone provided UVB can reach the skin. The synthesis of vitamin D_3 in the skin involves two stages: the photochemical transformation of 7-dehydrocholesterol to previtamin D_3 by UVB, followed by thermal isomerization of the previtamin to vitamin D_3 (cholecalciferol). Variables influencing the formation of previtamin D_3 in the skin include skin pigmentation, the intensity of the solar ultraviolet light, and environmental factors such as clouds, smog, clothing and sunscreen use (Grant et al., 2016; Wacker and Holick, 2013). Over-exposure to UVB will not lead to excess skin synthesis of vitamin D, as previtamin D as well as vitamin D_3 are irreversibly converted to inactive metabolites(Wacker and Holick, 2013). In the absence of UVB exposure, the requirement for vitamin D must be met from dietary sources. Mostly, the requirement is met partially or fully by diet, even in equatorial regions because people may embrace an urban or sun-avoiding lifestyle.

18b.2.2 Vitamin D_2 (ergocalciferol)

Vitamin D_2 (ergocalciferol) cannot be made by animals. The only source is ingestion of certain foods in the fungi kingdom. It is important to note the fungi kingdom is separate from the other eukaryotic life kingdoms of plants and animals and texts may erroneously refer to vitamin D_2 as a “plant” source. Sun-exposed mushrooms naturally provide vitamin D_2, while UV-exposed yeast and mushrooms added to the food supply enhance the vitamin D content of foods (Wacker and Holick, 2013).

The major metabolic steps involved in the metabolism of vitamin D_2 are similar to those for the metabolism of vitamin D_3 (cholecalciferol). The evidence about efficacy of vitamin D_2 versus vitamin D_3 suggests that although vitamin D_3 is the more active (Logan et al., 2013), vitamin D_2 is a reasonable alternative (Wacker and Holick, 2013). Hence, in the following discussion and in figure 18b.2, the term “vitamin ” refers to either or both vitamin D_2 and vitamin D_3 and their metabolites.

18b.2.3 Production of 25-hydroxyvitamin D

Vitamin D enters the circulation from the skin or from the lymph via the thoracic duct, bound to a specific vitamin D-binding protein. Vitamin D is transported to adipose tissue where it is stored, or to the liver, where it is hydroxylated to 25-hydroxyvitamin D ($25(OH)_2D$ [also called calcidiol when referring to $25(OH)_2D_3$], the major circulating form of vitamin D (Figure 18b.2).

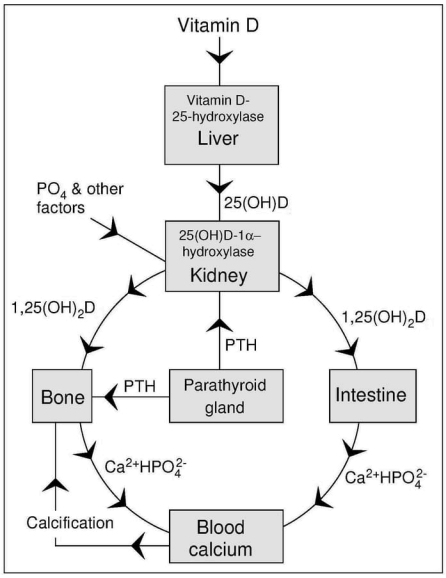


Figure 18b.2 Formation of vitamin D metabolites. Once the transport form 25(OH)D is made, the Endocrine pathway (section 18b.2.4) for synthesis of the active metabolite $1,25(OH)_2D$ is illustrated. Parathyroid hormone (PTH) directs synthesis of $1,25(OH)_2D$ in response to a need for calcium or phosphate. Plasma $1,25(OH)_2D$ stimulates intestinal calcium transport and bone calcium mobilization, then blunts PTH synthesis to turn cycle off (section 18b.13). Adapted from Holick, Kidney International 32: 912–929, 1987.

The plasma $25(OH)_2D$ metabolite has several fates. When there is a need for calcium (described below) then $25(OH)_2D$ is converted in the kidney by an enzyme ($25(OH)_2D$-1-α-hydroxylase) to produce the biologically active metabolite 1,25-dihydroxyvitamin D [$1,25(OH)_2D$] also called calcitriol. Circulating $25(OH)_2D$ can also be taken up by tissues and subsequently converted to $1,25(OH)_2D$ inside cells if the 1-hydroxylase enzyme has been activated. Finally, in the pathway for inactivation and excretion of vitamin D in bile, $25(OH)_2D$ is converted to 24,25-dihydroxyvitamin D by the enzyme 25-hydroxyvitamin D-24-hydroxylase (Wacker and Holick, 2013). The enzyme 25-hydroxyvitamin D-24-hydroxylase also inactivates $1,25(OH)_2D$ (calcitriol).

18b.2.4 Renal production of circulating 1,25-dihydroxyvitamin D (calcitriol)

The active form of vitamin D, $1,25(OH)_2D$ (calcitriol) that circulates in plasma is made in the kidney in the Endocrine Pathway as shown in Figure 18b.2. This synthesis of $1,25(OH)_2D$ is homeostatically controlled, mainly by the action of parathyroid hormone (PTH) in response to serum calcium levels, and fibroblast growth factor 23 (FGF-23) related to serum phosphate levels (Wacker and Holick, 2013), that regulate the activity of renal ($25(OH)D$-1-α-hydroxylase). For example, a decrease in plasma calcium prompts an increase in parathyroid hormone secretion from the parathyroid gland that acts to mobilize calcium stores from the bone. Parathyroid hormone also promotes the synthesis of $1,25(OH)_2D$ in the kidney which, in turn, stimulates the mobilization of calcium from the bone and increased intestinal calcium absorption (Figure 18b.2). Once plasma calcium levels are normal, the need for circulating $1,25(OH)_2D$ diminishes and there is no stimulation of the enzyme $25(OH)D$-1-α-hydroxylase. Thus, circulating levels of $1,25(OH)_2D$ are not related to vitamin D status as in deficiency low levels of $1,25(OH)_2D$ may reflect lack of the precursor metabolite $25(OH)D$. However, as vitamin D deficiency leads to secondary hyperparathyroidism with PTH-enhanced $1,25(OH)_2D$ production, vitamin D deficiency can be associated with normal to high $1,25(OH)_2D$ (calcitriol) levels.

Another reason why the level of $1,25(OH)_2D$ is not useful in assessing vitamin D status is because $1,25(OH)_2D$ is a very short-lived metabolite causing its own destruction by rapidly inducing synthesis of the enzyme 25-hydroxyvitamin D-24-hydroxylase (Wacker and Holick, 2013).

18b.2.5 Extrarenal production of 1,25-dihydroxyvitamin D (calcitriol)

The extrarenal pathway of $1,25(OH)_2D$ (calcitriol) is locally produced in almost every tissue in the body (Norman, 2008). As $1,25(OH)_2D$ acts locally, this synthesis pathway is called Paracrine / Autocrine. Activity of extra-renal 25(OH)D-1-α-hydroxylase is not regulated by the hormones that control renal 25(OH)D-1-α-hydroxylase (i.e., PTH and FGF-23). The activity of the enzyme must be induced in the cell. The diverse actions of $1,25(OH)_2D$, when acting locally as a transcription factor in many different cell types, are called “non-calcemic” or “non-skeletal” and include immuno-modulatory and cell-differentiating properties. It is these properties that have led researchers to investigate vitamin D and its derivatives in the pathogenesis of cancer, respiratory diseases, and immune responses. For further details of these noncalcemic functions see (Norman, 2008) and (Wacker and Holick, 2013).

18b.2.6 Serum 25(OH)D

Serum 25(OH)D is a biomarker of vitamin D exposure and status. Both metabolites 25(OH)D and $1,25(OH)_2D$ circulate in plasma. The former, 25(OH)D, reflects the sum of vitamin D from dietary intake and sunlight exposure, whereas plasma $1,25(OH)_2D$ concentrations reflect the immediate physiological need and are under homeostatic control in the kidney. Concentrations of $1,25(OH)_2D$ in plasma are about 0.1% of those of 25(OH)D. In vitamin D deficiency, serum $1,25(OH)_2D$ levels may be normal or even elevated, as a result of increased renal production of $1,25(OH)_2D$ in response to the rise in serum parathyroid levels (Wacker and Holick, 2013). In contrast, plasma 25(OH)D concentrations remain low until a reserve accumulates. As a result, the plasma 25(OH)D concentration reflects medium to long-term vitamin D availability from both dietary and endogenous sources, thus making it the best biomarker of vitamin D exposure and status.

This page titled 16.2: Metabolism of vitamin D (18b.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.3: Vitamin D deficiency in humans (18b.3)

Osteomalacia may occur in severely vitamin D deficient adults, a condition characterized by a failure in the mineralization of the organic matrix of bone. This results in weak bones, diffuse skeletal bone tenderness, proximal muscle weakness, and an increased frequency of fractures. Such disturbances are associated with serum 25(OH)D concentrations below 7.5nmol/L (Haddad and Stamp, 1974). After treatment with vitamin D supplements, serum 25(OH)D values rise and radiological lesions heal (Preece et al., 1975). Osteomalacia prevalence may be high globally due to lack of sun exposure, but remains largely undiagnosed due to the need to have radiographic data (Uday and Högler, 2019). It may occur in adults living in the tropics,who have no sun exposure, such as garment factory workers in Bangladesh (Islam et al., 2008). Low dietary intake may also play a role. A study in Germany of mainly white adults who died accidently, found the prevalence of osteomalacia to be at least 25% based on osteoid volume/bone volume ratio (Priemel et al., 2010).

Some adult patients with chronic renal failure, gastrectomy, intestinal malabsorption and steatorrhea arising from celiac disease, inflammatory bowel disease, pancreatic insufficiency, or massive bowel resection, may also develop osteomalacia. Functional disturbances have been described in adults with low serum 25(OH)D concentrations. These include secondary hyperparathyroidism, an increased bone turnover, and reduced bone mass (Chapuy et al., 1997). In the elderly, suboptimal vitamin D status decreases absorption of calcium, a factor associated with a lowering of the bone mineral content during postmenopausal aging.

Rickets occurs in infants and children with severe vitamin D deficiency. In rickets, abnormal softness of the skull (craniotabes) occurs. This may be accompanied by enlargement of the epiphyses of the long bones and of the costochondral junction (rachitic rosary). Bowlegs and knock knees may arise from these bone deformities. Rickets, arising from primary vitamin D deficiency may occur in infants in industrialized countries who are breast-fed without vitamin D supplementation. A supplement of 400 IU (10µg) per day to prevent rickets is often recommended for all infants from birth to 12mo of age, independent of their mode of feeding (Wagner and Greer, 2008; Munns et al., 2016). However, for breastfed infants whose mothers have an adequate vitamin D status, the content of vitamin D in breastmilk can be sufficient because the vitamin D content in breastmilk is dependent on maternal status (Stoutjesdijk et al., 2017). Nutritional rickets can occur in older children, particularly during the adolescent growth spurt (Beck-Nielsen et al., 2009; Uday and Högler, 2019). However, some nutritional rickets is also caused by a lack of calcium, so both vitamin D and calcium should be monitored (Munns et al., 2016; Uday and Högler, 2019).

Metabolic defects also cause rickets, including both vitamin D-resistant rickets (familial hypophosphatemia) and vitamin D-dependent rickets (VDDR type 1). The latter condition is a deficiency of the 25(OH)$_2$D-1-hydroxylase enzyme, while vitamin D-resistant rickets is a defect in proximal renal tubular resorption of phosphate. The yearly incidence of hypophosphatemic rickets in infants 0–0.9y is about 3.9 per 100,000: vitamin D-resistant rickets (VDDR type 1) is very rare (Beck-Nielsen et al., 2009).

This page titled 16.3: Vitamin D deficiency in humans (18b.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.4: Food sources and dietary intakes (18b.4)

Natural food sources rich in vitamin D are restricted to fatty fish, organ meats, and UV-exposed mushrooms. In some countries — for example, the United States, Canada, and Finland — fluid milk is fortified with vitamin D. In the United Kingdom and Europe, low amounts of vitamin D can be added to some breakfast cereals, margarine, fat spreads, and vegetable oils, breakfast beverages, and breads (Calvo et al., 2005). In low and medium income countries, fortification of wheat, edible plant-based oil or milk, could reduce vitamin D deficiencies (Cashman and O'Dea, 2019).

Intakes from national nutrition surveys provide data on vitamin D in different countries. They show low (3.0µg/d) vitamin D intakes on average in countries where there is little or no vitamin D fortification, such as the U.K. However in countries such as Japan, fatty fish may be consumed in sufficient quantities to provide vitamin D intakes > 7µg/d (Calvo et al., 2005). Canada, where there is mandatory fortification of milk and margarine, has an average intake of 5µg/d which does not meet the adult US/Canadian Estimated Average Requirement (EAR) of 10µg/d. The Canadian government will double the level of mandatory fortification by 2023 to try to reduce the prevalence of inadequacy (Vatanparast et al., 2020).

High Intakes following self-dosing with excessive amounts of vitamin D supplements have been described, although the dose required to induce vitamin D toxicity is uncertain (Hathcock et al., 2007). Vitamin D intoxication does not arise from the consumption of conventional foods (including fortified foods), nor does excess exposure to UVB through sun or artificial lamps cause toxicity (Marcinowska-Suchowierska et al., 2019). Cases of toxicity have arisen, however, from accidental overfortification of milk with vitamin D_3, from uncontrolled use of vitamin D mega-doses, and from inappropriate use of vitamin D metabolites. An increased risk for vitamin D toxicity is also associated with certain diseases, including sarcoidosis, tuberculosis, and genetic disorders of rare polymorphisms of enzymes involved in vitamin D metabolism (e.g., idiopathic infantile hypercalcemia)(Wacker and Holick, 2013). Signs of vitamin D toxicity include hypercalcemia (i.e., elevated serum calcium concentrations) and hypercalciuria (elevated urine calcium levels). Hypercalcemia arises from hyperabsorption of intestinal calcium and, to a lesser degree, from the release of calcium from bone and can lead to calcification of soft tissues such as arteries (arteriosclerosis) and the kidney (nephrocalcinosis). The hypercalciuria associated with vitamin D toxicity reflects the presence of excess calcium in the serum arising from the release of calcium from bone.

This page titled 16.4: Food sources and dietary intakes (18b.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.5: Nutrient reference values for vitamin D (18b.5)

Unlike other essential nutrients, measurements of vitamin D intake using traditional dietary assessment methods cannot be used to measure dietary exposure because some of the requirement for vitamin D can be met from skin synthesis, as noted earlier. Instead a surrogate biomarker is used to measure exposure to vitamin D; serum 25 hydroxyvitamin D (25(OH)D) is the biomarker of choice. Serum 25(OH)D measures exposure to vitamin D from the effects of both diet and sunlight as noted earlier, and associations between concentrations of serum 25(OH)D and functional biomarkers of bone health have been considered when setting Nutrient Reference Values (NRVs) for vitamin D by several agencies. The evidence for extraskeletal outcomes has been considered inadequate, inconsistent, or insufficient to develop Nutrient Reference Values (NRVs). Nevertheless, in view of the variation in sunlight exposure and the variable response to that exposure, as well as concerns about skin cancer, several agencies, including the IOM(2011), EFSA (2016), and the UK (2016) have set their NRVs based on the assumption of minimal or no sunlight exposure. This means that in the presence of cutaneous vitamin D synthesis, the requirement for dietary vitamin D may be lower than that set, or may even be zero.

The US IOM (2011), as an example, set an Estimated Average requirement (EAR), the requirement to satisfy the need of half the population, of 10µg/d for all children and adults age 1–70y, as necessary to maintain bone health and achieve a serum 25(OH)D concentration of 40nmol/L (16ng/mL). The Recommended Dietary Allowance (RDA), designed to cover the requirements for 97.5% of the population, was set at 15µg/d, a level that corresponded to a serum 25(OH)D concentration of 50nmol/g (20ng/mL). A higher RDA level (i.e., 20µg/d) was recommended for those over 70y, due to age-related inefficiencies in vitamin D metabolism. For infants 0–12months, IOM set an Adequate Intake (AI) of 10µg/d because there was not enough information to establish an EAR for this group.

Several other countries and regions have also restricted their recommendation to a single level — the AI (or equivalent) — because of insufficient evidence, including Australia and New Zealand (NHMRC, 2006), the European Union (EFSA, 2016), and most recently, the United Kingdom (SACN, 2016), although here only for infants and children less than 4y. However, the target serum 25(OH)D concentrations on which the AIs are based vary, ranging from at least 25nmol/L (10ng/mL) in Australia and New Zealand (NHMRC, 2006), and the UK (SACN, 2016), to 50nmol/L (20ng/mL) in the EU(EFSA, 2016). As a consequence, the AI levels range from about 10µg/d to 15µg/d for most age groups, similar to those for several other European countries (Spiro and Buttriss, 2014).

The UK , like the US has also set a Recommended Nutrient Intake (equivalent to the RDA) for all life-stage groups (including pregnant women, lactating women, and the elderly), except those < 4y. However, for the UK (SACN, 2016), the RNI is 10µg/d in contrast to the 15µg/d RDA set by IOM(2011). No EAR has been set by the UK. Cashman et al. (2008), have emphasized that dietary levels of about 10–15µg/d may not be adequate to keep most of the adult population in Europe above a higher target value of 50nmol/L in winter without adequate sun exposure in the summer season.`

Several agencies have established a Tolerable Upper Intake Level (UL) for vitamin D to discourage potentially dangerous self-medication. The UL is defined as the highest average intake that is likely to pose no risk. For infants 0–6mo, the IOM set an UL of 25µg/d, rising with age through childhood to 100µg/d for persons aged 9y and older. This upper limit corresponds to an average serum total 25(OH)D levels of 125nmol/L (50ng/mL) (2011).

The UL values for the European Food Safety Authority (EFSA) start for infants aged 7–11mo with an UL of 35µg/d which rises to 100µg/d for persons aged > 11y (EFSA, 2016). Data on the long-term adverse effects of high doses of vitamin D are limited, however, and caution is necessary when setting an UL for chronic intake (Aloia, 2011).

The Endocrine Society, focusing on clinical use for patients at high risk of deficiency with rickets, osteomalacia, osteoporosis, chronic kidney disease, etc., and some special populations (e.g., pregnant women) , have set a higher adult UL of 250µg/d (Holick et al., 2011) , a recommendation that is not appropriate for healthy individuals in the population. This distinction is important, and if not recognized, could result in inappropriate dietary recommendations for healthy individuals, for whom the NRVs should be used (Aloia, 2011).

This page titled 16.5: Nutrient reference values for vitamin D (18b.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.6: Biomarkers of vitamin D status (18b.6)

Historically, vitamin D status was assessed indirectly by measuring alkaline phosphatase activity, as well as calcium and phosphorus concentrations in serum: all very nonspecific indices. Methods are now available for the direct measurement of vitamin D metabolites in serum, and these are described below. If possible, these measurements should be performed in conjunction with an assay of serum parathyroid hormone and some functional assessment of skeletal health. In adults, this assessment may include measurement of bone mineral content or bone mineral density. In children, in extreme cases of rickets, bony deformities such as enlarged fontanelle, rachitic rosary, and swollen joints are clinical signs of rickets, whereas knock knees or bowed legs are clinical signs of the associated osteomalacia noted in growing children (Uday and Högler, 2019).

18b.6.1 Serum 25-hydroxyvitamin D

Serum 25-hydroxyvitamin D is the circulating metabolite of vitamin D that is the most abundant and has the longest half-life of all the vitamin D derivatives. Concentrations of 25(OH)D in serum (or plasma) are also the most useful measure of vitamin D exposure and status in humans, as they reflect the total supply of vitamin D from both cutaneous synthesis and dietary intake of either vitamin D_2 or vitamin D_3 (Wacker and Holick, 2013).Moreover, they can be used to define vitamin D deficiency, insufficiency, hypovitaminosis, sufficiency, and toxicity (IOM, 2011). Concentrations in healthy adults vary from 30–130nmol/L, depending in part on exposure to solar ultraviolet light.

18b.6.2 Factors affecting serum 25 hydroxyvitamin D

Seasonal and latitude effects on serum 25 hydroxyvitamin D status are marked in many areas of the world. Subjects living north of latitude 33°N and south of latitude 33°S (Wacker and Holick, 2013) have reduced dermal vitamin D synthesis during winter, and show the highest serum 25(OH)D levels in the late summer, and the lowest in late winter. In a U.K. national survey of people > 65y (Finch et al., 1998), mean plasma 25(OH)D concentrations were significantly higher in free-living participants surveyed in the summer (July–September) than in the winter (Figure 18b.3). Finch et al. (1998) further noted that in the summer, only 6% of a free-living group of elderly subjects had plasma 25(OH)D concentrations < 25nmol/L compared to 35% of those who were institutionalized.

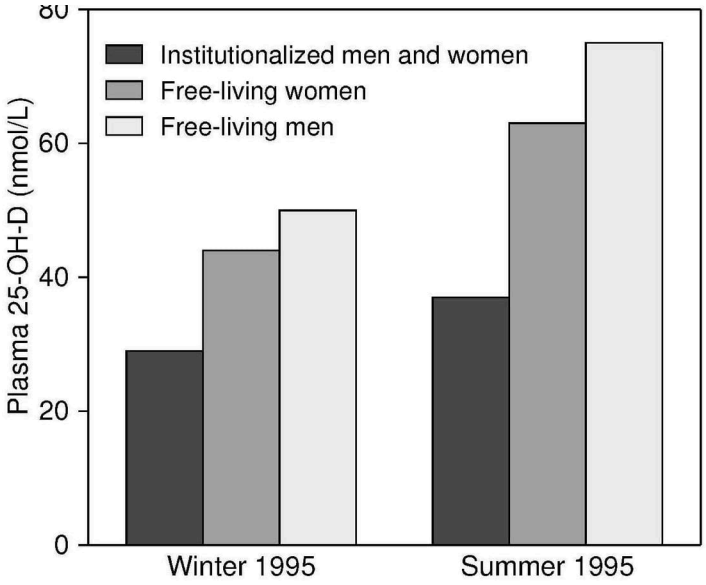


Figure 18b.3. Comparison of mean plasma 25(OH)D levels by season for free-living elderly men and women and participants in institutions. Data from Finch et al., 1998, National Diet and Nutrition Survey: People Aged 65 Years or Over. The Stationery Office, London.

Only UVB rays (290–315nm) elicit synthesis of vitamin D_3 (cholecalciferol). When the incident angle of the sun is low, UVB does not reach the earth (Grant et al., 2016). The "Shadow Rule" states that if one's height is longer than the length of one's shadow, vitamin D synthesis is possible; if not, the sun's incident angle is too low to provide UVB. This also explains why the best time for vitamin D synthesis is between 10:00 and 14:00 in summer in temperate countries or year-round in low latitude regions. Other factors affecting dermal synthesis are described below.

Age-related changes from infancy to adulthood in the concentrations of serum 25(OH)D can be marked. Newborn infants have serum concentrations that correlate with maternal 25(OH)D concentrations. Breastfed infants, as noted above, are at risk of poor vitamin D status unless mothers have an adequate vitamin D status and can pass on metabolites in their milk (Stoutjesdijk et al., 2017). Vitamin D metabolites are lower when measured in cord blood. Serum 25(OH)D levels in children of both sexes decline with increasing age (Gregory et al., 2000), a trend that may reflect both behavioral factors (diet and sun exposure) as well as biological effects (increasing requirements of vitamin D with growth).

Studies indicate that calcium absorption by premature infants is not only vitamin D dependent; however, concern remains about the lack of data to outline the vitamin D needs for premature infants (Taylor et al., 2019).

Older adults are particularly vulnerable to low levels of serum 25(OH)D (McKenna et al., 1985). Lifestyle factors that reduce vitamin D status include low dietary intakes of vitamin D and limited sun exposure with a greater use of sunscreens and umbrellas. In addition, there are biological reasons for the low status including a reduced capacity of the skin to produce vitamin D resulting from a reduction of 7-dehydrocholesterol in the skin, and impaired intestinal absorption of ingested vitamin D (Wacker and Holick, 2013). In countries such as Canada where low vitamin D status is related to osteoporosis, it is recommended that older adults take vitamin D supplements and age-related declines in vitamin D status as measured by serum 25(OH)D have not been observed in Canadian national survey data (Brooks et al., 2017).

Sex and Gender differences in concentrations of serum 25(OH)D have been noted, although no consistent pattern has emerged from national survey data in the USA (Wacker and Holick, 2013) and Canada (Brooks et al., 2017). Differentiating biological (sex) effects from lifestyle effects (gender)such as differences in food preferences, clothing, and amounts of sun exposure during work or leisure may be difficult. In some countries, consumption of fortified foods and supplements as well as sunscreen use and sun avoidance practices, differ between males and females of all ages.

Skin pigmentation differences, often seen when comparing ethnic or racial groups, greatly influences serum 25(OH)D concentrations. Skin pigmentation reflects the amount of the pigment melanin that absorbs UVB rays and lowers the amount of UVB acting on previtamin D. In the USA, for example, serum 25(OH)D concentrations in African-Americans and Hispanics are much lower than in non-Hispanic whites (2011). In European countries, African immigrants now living in northern latitudes have a greater

risk of vitamin D deficiency compared to non-migrants. As an example, of the patients with nutritional rickets in Denmark, 74% were immigrant children (Beck-Nielsen et al., 2009). Not all differences in vitamin D status between ethnic groups are due to skin pigmentation. Behavioral differences such as dietary intakes, clothing preferences, and sun avoidance practices are also important factors.

Melanin in skin does not block all cholecalciferol synthesis, but it is slowed. Webb and Engelsen (2006) have calculated the time needed for a person with each of the different Fitzpatrick categories of skin type to burn with sun exposure. Also shown in Table 18b.1, are the times needed to synthesize 1000 IU of vitamin D. The times are taken from a more complex analysis that depends on time of day and day of the year in areas that experience marked seasons, and the amount of skin exposed to sun. As shown, all skin types can synthesize vitamin D, but the time needed is longer as skin pigmentation increases (higher Fitzpatrick numbers). The times in the southern hemisphere will be influenced by variation in theozone layer thickness.

Table 18b.1 Association of Fitzpatrick skin types with capacity of skin for vitamin D synthesis at 42°N on 21 June at 10:30 hours while synthesis in spring or fall equinox or a higher latitude (62.5°N) would be double. Source Webb and Engelsen (2006) Photochemistry

Fitzpatrick Skin Type	How skin responds to sun exposure
I	Always burn never tan
II	Burn slightly then tan slightly
III	Rarely burn tan moderatelu
IV	Never burn, tan moderately e.g. Mediterranean
V	Never burn, tan darkly Asian, Indigenous American, Pacific Islander
VI	Never burn, tan very darkly; Australian Aborigine, Tamil, West African

Sunscreen lotions are used to deliberately block UV rays reaching the skin. Sunscreens are labeled with a sun protection factor (SPF) number which indicates the amount of UV blocked. An SPF blocks at 1/SPF, so that a product having an SPF of 8 would allow only 1/8 (12.5%) of the UV to penetrate the lotion and reach the skin. Theoretically sunscreens should reduce vitamin D synthesis. In a national survey in Canada, however, participants answering "yes" to using sunscreen had *higher* 25(OH)D levels (2.4 ±1.1nmol/L; $P < 0.001$) (Brooks et al., 2017). Several reasons may account for this seemingly abherrant finding. Users may apply sunscreens poorly or incompletely (for example answer "yes" to use but only apply to face). Alternatively, sunscreen users may spend more time outdoors. Other behavioral differences may exist, for example in the Canadian survey, sunscreen users were more likely to take vitamin D supplements than nonusers.

Smoking is associated with lower serum 25(OH)D concentrations. This may partly explain the reported increased risk of osteoporosis among smokers. The mechanism is unclear, but the relationship does not appear to result exclusively from additional confounding lifestyle factors (Brot et al., 1999). In a large study of adults in Australia, both male and female nonsmokers, including ex-smokers, had higher mean levels of serum 25(OH)D compared to current smokers (Gill et al., 2017).

Obesity, prevalent in many population groups worldwide, is associated with a trend towards lower serum 25(OH)D levels (Brooks et al., 2017; Wacker and Holick, 2013). Biologically, this trend can be attributed to vitamin D, whether from cutaneous or dietary sources, being deposited in adipose tissue, where it is not bioavailable (Wacker and Holick, 2013). Some researchers have found that obesity in men has less of an effect on reducing 25(OH)D than in women, perhaps because of gender differences in behavioral aspects such as sun avoidance (Rockell et al., 2006). Nevertheless, body weight must be considered when evaluating vitamin D status

Disease conditions affecting the gastrointestinal tract, the liver and kidneys may cause a secondary deficiency of vitamin D (Table 18b.2). Diseases causing fat malabsorption will reduce the absorption of dietary vitamin D, which could be significant in winter. Liver disease will prevent the conversion of vitamin D to 25(OH)D, whereas kidney disease prevents the conversion of 25(OH)D to the active form (i.e. $1,25(OH)_2D$, calcitriol) in the endocrine pathway. Disease states such as intestinal malabsorption and steatorrhea caused by pancreatic insufficiency, inflammatory bowel disease, celiac disease, or massive bowel resection have also been associated with lower serum 25(OH)D concentrations. Here, vitamin D depletion arises from malabsorption of dietary vitamin D.

Table 18b.2 Secondary causes of vitamin D deficiency. Source Holick (2007). Vitamin D deficiency. The New England Journal of Med 357(3), 266–281.

Causes of Secondary vitamin D Deficiency	
Pathology	**Diseases**
Malabsorption of fat reduces absorption of dietary vitamin D	Cystic fibrosis Celiac disease, Whipple's disease, Crohn's disease, Bypass surgery.
Liver failure prevents production of 25(OH)D	Cirrhosis Hepatitis
Inability to produce 1,25(OH)2D in kidney	Chronic kidney disease
Medications	
Drugs reducing Vitamin D absorption	Cholesterol-lowering agents: cholestyramine Weight loss drug orlistat and food additive olestra
Drugs reducing 25(OH)D levels due to increased catabolism	Anticonvulsant medications such as carbamazepine, phenobarbital, and phenytoin, gabapentin. Antiretrovirals agents such as ritonavir and efavirenz, valproic acid (AIDS treatment) Histamine H2 receptor antagonist cimetidine
Drugs Impairing vitamin D metabolism	Oral corticosteroids such as glucocorticoids

Medication use can affect vitamin D status. Any drug which affects liver or kidney cytochrome enzymes will likely affect conversions of vitamin D metabolites. Table 18b.2 provides a list of drugs known to impact vitamin D status. While this list does not cover all possible secondary causes of deficiency, it emphasizes the need to monitor disease states and medication use as possible reasons for vitamin D deficiency. As is described below, the requirement for vitamin D may be elevated in persons who have chronic conditions or for whom medication use is required.

Magnesium status may impact 25(OH)D levels and therefore vitamin D status through the requirement for two enzymes of vitamin D metabolism: 25(OH)D-1-α-hydroxylase and 25(OH)D-24-hydroxylase. In magnesium deficiency, there is a reduction in the active form $1,25(OH)_2D$ associated with “Mg-dependent vitamin-D-resistant rickets” (Dai et al., 2017). More research is needed to determine the intake of magnesium that affects vitamin D status.

Analytical methods have a marked effect on serum 25(OH)D concentrations. To overcome inter-assay differences, and establish the accuracy and precision of the assay, verified standards should be run with every batch using the Vitamin D Standardization Program. See Section 18b.9.

This page titled 16.6: Biomarkers of vitamin D status (18b.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.7: Interpretive criteria (18b.8)

Table 18b.3 shows a lack of consensus in defining cutoff points for serum total 25(OH)D (nmol/L) (i.e., the sum of serum 25(OH)D_2 and 25(OH)D_3 concentrations). The uncertainty is attributed in part to the lack of standardization of the laboratory measurement of serum total 25(OH)D and the lack of potential functional biomarkers of vitamin D status. Only three of the numerous meta-analyses completed to date are based on standardized serum total 25(OH)D levels (Sempos and Binkley, 2020).

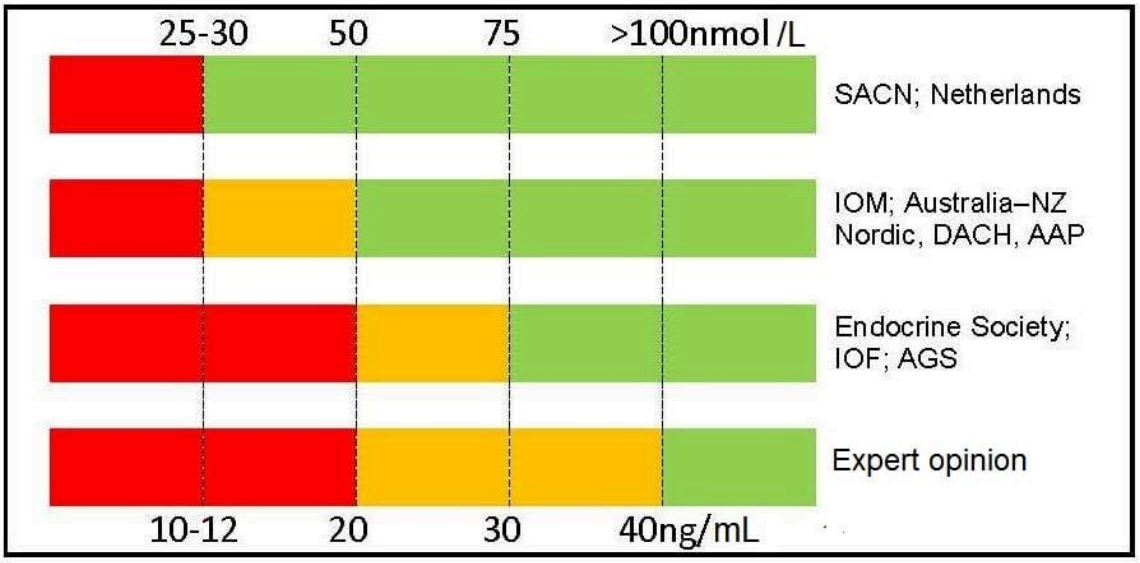


Table 18b.3. Recommendations for interpreting serum levels of 25-hydroxyvitamin D. Colour code: red denotes a state of severe deficiency (danger) that has to be corrected without exception; orange denotes a state of mild deficiency (modest concern), in which intervention is desirable; green denotes a state of sufficient supply that does not benefit from additional supplementation. AAP, American Academy of Pediatrics; AGS, American Geriatrics Society; DACH, Deutschland, Austria and Confoederatio Helvetica; IOF, International Osteoporosis Foundation; IOM, Institute of Medicine; SACN, Scientific Advisory Committee on Nutrition. Modified from Bouillon (2017)

Agreement exists Table 18b.3 between two groups of nations (i.e., United Kingdom, the Netherlands with Australia/New Zealand, European Union, USA) for the cutoff point to define severe deficiency (i.e., 25–30nmol/L, 10–12ng/mL), but there is much less agreement for cutoff points to define vitamin D insufficiency, vitamin D sufficiency, and vitamin D toxicity. Conflicting recommendations have been compiled by some non-governmental medical societies and organizations.

In 2011, the US Institute of Medicine (IOM) published cut-off points for serum total 25(OH)D that represent public health guidelines for generally healthy non-diseased populations Table 18b.4

Table 18b.4. Comparison of Institute of Medicine and Endocrine Society cut-points for serum total 25-hydroxyvitamin D (ng/mL) [nm

Interpretation	IOM (ng/mL)	Endocine Society (ng/mL)
Deficient	<12	<20
Insufficient	12–20	21–29
Sufficient	20–30	30–100
No added benefit	30–50	
Possible harm	>50	>100

(IOM, 2011). A summary report is available in Ross et al. (2011). The lower limit of adequacy, interpreted as deficiency (i.e.,< 12ng/mL; 30nmol/L), was based on relationships to biomarkers of bone health. Cutoff points for insufficiency, sufficiency, no added benefit, and possible harm were also defined by IOM, and have been adopted by several other organizations, including the Global Consensus Recommendations on the Prevention and Management of Nutritional Rickets (Munns et al., 2016, Sempos and Binkley, 2020).

Note the levels set by IOM differ from those set for clinical practice by the Endocrine Society in 2011, and presented in Table 18b.4 (Holick et al., 2011).However, these are intended as guidance for clinicians for the evaluation, treatment, and prevention of vitamin D deficiency, with emphasis on the care of patients who are at high risk of deficiency (e.g., those with rickets, osteomalacia, osteoporosis, chronic kidney disease, etc). Such practice guidelines to treat disease should not be applied to the apparently healthy population. Clearly, more research is urgently required using assays that meet the standardized criteria developed by the Vitamin D Standardization Program (VDSP) (Section 18b.9) to develop a consensus on a single set of rigorous interpretive criteria for serum total 25(OH)D levels relative to vitamin D status and interventions that are appropriate for both public health and clinical needs. In the meantime, readers are advised to use the IOM (2011) cutoffs proposed and listed in Table 18b.4 for public health use.

This page titled 16.7: Interpretive criteria (18b.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)

There are three main methods for measuring serum 25(OH)D: immunoassay, high performance liquid chromatography (HPLC), and liquid chromatography with tandem mass spectrometry (LC-MS/MS) However, there are two main steps that must be undertaken before selecting a method for assaying total 25(OH)D in serum. The first step is verification of “fit-for-purpose” . This step is only necessary when an immunoassay is the intended method and is necessary because some immunoassays are not appropriate for certain patient population groups (e.g., those with certain disease states, pregnant women, and vegetarians such as vegans) (Sempos and Binkley, 2020). In addition, the chosen immunoassay should have an appropriate measurement range for the intended study population (e.g., measure 25(OH)D levels in persons who are deficient). The “fit for purpose” step can be accomplished by testing the immunoassay against a Vitamin D Standardization Program (VDSP) standardized LC-MS/MS assay.

The second step required is standardization, a process whereby the intended assay is calibrated to meet the VDSP performance criteria (i.e., total coefficient of variation (CV) < 10% and a mean bias with the range of –5% to +5%). Some assays, even when based on a similar methodology, are less accurate and precise, making comparison of data across assays or laboratories difficult. The gold standard analytical method is LC-MS/MS and that should always be the method of choice for national surveys. However, even the LC-MS/MS method must meet the assay standardization criteria of the VDSP. Serum total 25(OH)D measurements can be 'prospectively' standardized or 'retrospectively' standardized, using methods developed by VDSP; see Sempos et al., 2017 and Durazo-Arivizu et al., 2017 for more details. A standardized laboratory measurement is defined as one that provides the 'true' serum total serum 25(OH)D concentrations. Several NIST Standard Reference Materials — SRM 972a, 2973, and 1949 — which provide target values for $25(OH)D_2$ and $25(OH)D_3$ are available to check on the accuracy and precision of the chosen assay. NIST SRM 1949 also provides target values for serum total 25(OH)D values for the first, second, and third trimester of pregnancy and these should be used in studies of pregnancy.

This page titled 16.8: Measurement of serum 25-hydroxyvitamin D (18b.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)

The active form of vitamin D is $1,25(OH)_2D$ (also called calcitriol) interacts with its nuclear receptor in the intestine, bone, and kidney for most of its functions to regulate calcium and bone metabolism. $1,25(OH)_2D$ also has many other noncalcemic cellular actions that reside in the Paracrine/Autocrine pathway (Norman, 2008; Wacker and Holick, 2013).

$1,25(OH)_2D$ in serum is not a useful marker of vitamin D status because it has a short half-life (4–6h) and levels are under stringent homeostatic regulation by factors at the site of $1,25(OH)_2D$ synthesis in the kidney. Hence, it is not surprising that no seasonal variation in serum $1,25(OH)_2D$ concentrations has been reported (Landin-Wilhelmsen et al.,1995).

Synthesis of $1,25(OH)_2D$ in the kidney is stimulated by low serum concentrations of calcium or phosphorus and is inhibited by excess $1,25(OH)_2D$. In cases of vitamin D sufficiency, a positive relationship exists between serum $1,25(OH)_2D$ and 25(OH)D concentrations (Need et al., 2000), presumably because 25(OH)D is the substrate for $1,25(OH)_2D$. In vitamin D deficiency, however, this relationship is reversed because with a fall in serum 25(OH)D concentrations there is a rise in the concentration of parathyroid hormone resulting in an increase in the renal production of $1,25(OH)_2D$ (Figure 18b.2) (Wacker and Holick, 2013). As a result, the circulating concentrations of $1,25(OH)_2D$ often become normal or even elevated. In contrast, serum $1,25(OH)_2D$ levels decrease in renal disease (which affects the enzyme, 25(OH)D-1-α-hydroxylase); levels are very low in anephric patients (i.e., those lacking a functional kidney), and in patients on hemodialysis. Even in normal healthy subjects, concentrations of serum $1,25(OH)_2D$ are in the picomolar range, making analysis difficult (Zittermann, 2003). Serum $1,25(OH)_2D$ concentrations ranging from 60–100pmol/L are normal in adults; concentrations in children tend to be higher. Methods of analysis include HPLC, LC-MS/MS and radioimmunoassay. (Zittermann et al., 2016).

This page titled 16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.10: Other vitamin D metabolites (18b.11)

Alkaline phosphatase (EC 3.1.1.1) activity in serum can be used as an indirect measure of vitamin D status. Activity increases in osteomalacia in adults and childhood rickets but is generally normal in osteoporosis. Increases in the activity of alkaline phosphatase are usually proportional to the severity of vitamin D depletion. For example, elderly Irish individuals with serum 25(OH)D levels indicative of severe or marginal vitamin D depletion had slightly higher serum alkaline phosphatase activity than those with 25(OH)D levels classified as replete (Figure 18b.4).

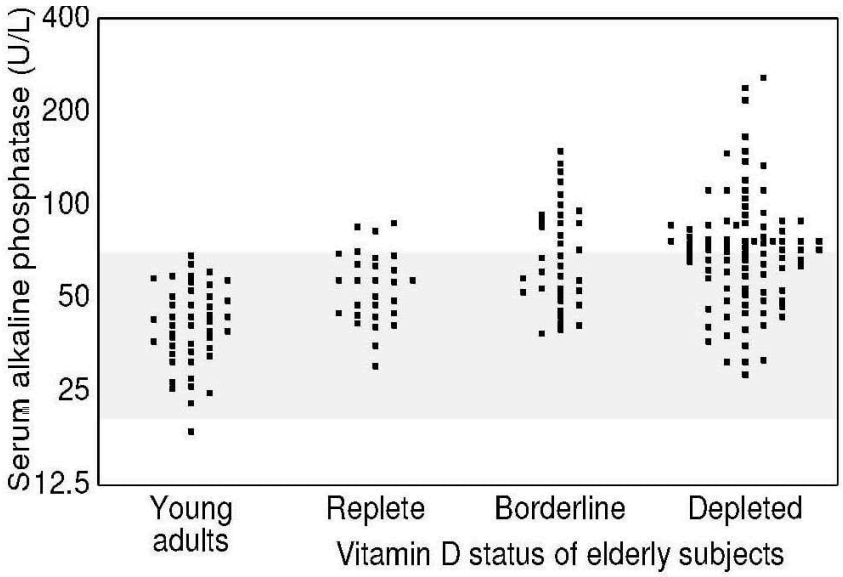


Figure 18b.4. Comparison of mean plasma 25(OH)D levels by season for free-living elderly men and women and participants in institutions. Data from Finch et al., 1998, National Diet and Nutrition Survey: People Aged 65 Years or Over. The Stationery Office, London.

Seasonal changes in serum alkaline phosphatase activity have also been observed in cross-sectional studies, with levels decreasing with seasonal rises in serum 25(OH)D levels (McKenna et al., 1985).

Serum alkaline phosphatase activity is also affected by sex and age, again in the opposite direction to changes in the levels of serum 25(OH)D. Serum alkaline phosphatase activity is significantly higher in females relative to males and in older versus younger adults; activity is also higher in growing children and pregnant women, especially during the third trimester (McKenna, 1992). In older surveys such as the U.K. national survey of young people 4–18y, mean alkaline phosphatase activity was measured. The lowest level was found in the oldest adolescents (15–18y), especially among the girls (Gregory et al., 2000).

The activity of serum alkaline phosphatase is also altered by various disease states such as hyperparathyroidism, Paget's disease, secondary bone cancer, and cholestasis (Sauberlich, 1999). Serum alkaline phosphatase activity may decrease in zinc deficiency but, as noted earlier, appears to be close to normal in osteoporosis in contrast to osteomalacia where it is high (Uday and Högler, 2019).

In general, measurement of alkaline phosphatase activity in serum is best used to confirm a clinical diagnosis of vitamin D deficiency, or as a screening tool, but it is not very useful for detecting subclinical vitamin D deficiency. The diagnosis of osteomalacia can be made in the presence of high alkaline phosphatase activity accompanied by high PTH, low dietary calcium intake (< 300mg/d) and/or low serum 25(OH)D (< 30nmol/L) (Uday and Högler, 2019).

18b.12.1 Interpretive criteria

Serum alkaline phosphatase activity is normally expressed as U/L. The reference range for normal adults is 30–135U/L (Gregory

et al., 2000). Total plasma alkaline phosphatase activity was measured in the past U.K. national surveys (Gregory et al., 2000, Finch et al., 1998) , with the exception of the survey on pre-school children. Mean, median, and lower and upper 2.5 or 5th percentiles by age and sex are presented.

18b.12.2 Measurement of alkaline phosphatase

Several methods are available for the assay of serum or plasma alkaline phosphatase (ALP); for plasma, heparinized blood samples should be used (Bessey et al., 1946). Total alkaline phosphatase measures all sources of enzyme activity including liver, while bone-specific alkaline phosphatase (BALP) measures activity derived from osteoblast activity. Higher total or bone alkaline phosphatase levels accompanied by low urinary calcium could prompt investigation of vitamin D deficiency (Kennel et al., 2010). The assay of total alkaline phosphatase activity is sufficient evidence as long as other liver enzymes are normal. The activity of serum or plasma alkaline phosphatase should be expressed as U/L. The within-subject coefficient of variation for serum/plasma alkaline phosphatase activity ranges from 4.6% to 9.2%, depending on the time frame and dietary regimen (Gallagher et al., 1989). The enzyme is reasonably stable in frozen serum or plasma.

This page titled 16.10: Other vitamin D metabolites (18b.11) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.11: Serum parathyroid hormone (18b.13)

Parathyroid hormone (PTH) levels in serum or plasma are considered to be a functional biomarker of vitamin D status in the normocalcemic state. In vitamin D deficiency, when calcium absorption is reduced, serum PTH levels rise to induce calcium mobilization from the bone and increase tubular reabsorption of calcium in the kidney. In this way, serum calcium is maintained at a physiologically optimum level. As a result, serum PTH concentrations are inversely related to serum 25(OH)D levels (Chapuy et al., 1997; Need et al., 2000; Wacker and Holick, 2013).

Threshold values for serum 25(OH)D concentrations that induce an increase in serum parathyroid secretion vary. In a study of Australian postmenopausal women, the threshold value was 40nmol/L (Need et al., 2000), whereas in French adults it was 78nmol/L (Chapuy et al., 1997). In the 2011 IOM report, studies examined showed that PTH is inversely associated with serum 25(OH)D concentrations at lower 25(OH)D concentrations but there was inconsistent evidence for a threshold above 27nmol/L. Such variable evidence for a threshold may result from different assays being used, both to measure serum PTH and serum 25(OH)D. Even small increases in PTH levels may have a negative influence on bone mass and increase the risk of non-vertebral fracture (Chapuy et al., 1997). Hence, a combination of serum PTH and serum 25(OH)D concentrations is often recommended as an indicator of vitamin D status.

Serum PTH levels increase with age, independent of 25(OH)D, ionized calcium, phosphate, and renal function (Carrivick et al., 2015). Further research is required to explore the underlying mechanisms and clinical relevance. Season also influences PTH concentrations, as might be expected from seasonal variations in 25(OH)D. Parathyroid hormone can be measured by radioimmunoassay using commercial kits. (Zittermann et al., 2016).

This page titled 16.11: Serum parathyroid hormone (18b.13) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

16.12: Calcium and phosphorus in serum and urine (18b.14)

Many studies of vitamin D status have included the measurement of calcium and phosphorus concentrations in serum or urine. The measurements are most useful if combined with the measurements of serum 25(OH)D and PTH concentrations. In vitamin D deficiency in infants and children, the serum calcium and phosphorus levels are usually reduced. For example, significantly lower mean serum calcium concentrations were reported in French neonates with serum 25(OH)D concentrations < 30nmol/L and elevated PTH concentrations, in comparison with individuals with normal values for these biochemical parameters (Zeghoud et al., 1997). Unlike total serum calcium alone, the existence of such a triad of biochemical disturbances strongly indicates vitamin D deficiency.

Serum calcium is also used to identify possible vitamin D intoxication. In such cases, concentrations of serum 25(OH)D and serum calcium are elevated and provide additional evidence for hypervitaminosis D.

The response of urinary calcium and phosphorus levels to changes in vitamin D status varies, as excretion of these minerals is also affected by dietary intakes. Early osteomalacia can be detected as a decrease in urinary calcium to creatinine ratio (Uday and Högler, 2019). Changes in urinary calcium and phosphorus concentrations, however, are not specific for vitamin D status. Details of the measurement of calcium and phosphorus in serum and urine are given in Chapter 23.

This page titled 16.12: Calcium and phosphorus in serum and urine (18b.14) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17: VITAMIN C (CHAPATER 19)

Abstract

Vitamin C (ascorbic acid) is involved in a variety of biological processes, acting primarily as an electron donating enzyme cofactor and reducing agent (antioxidant). Clinical deficiency of vitamin C results in scurvy, a relatively rare disease in high-income countries. Nevertheless, moderate vitamin C deficiency (hypovitaminosis C) has been described in selected vulnerable groups such as the institutionalized elderly, with gingival inflammation and fatigue appearing to be the most sensitive clinical markers for moderate vitamin C deficiency. Excess intakes of vitamin C may produce osmotic diarrhea in some people, and although oxaluria has been reported, this measurement is prone to errors in the preparative or investigative procedure. There are currently no reliable functional indices of vitamin C nutriture. Static biochemical indices such as plasma or leukocyte ascorbic acid concentrations are most frequently used. The former reflect recent dietary intake within a limited range (30–80mg/d). Plasma ascorbic acid concentrations increase linearly with dietary intake but only up to a plateau of about 70μmol/L, and thus are unsuitable for detecting excessive intakes of vitamin C. A number of non-nutritional factors, including sex, body weight, smoking, and severe infection, affect plasma ascorbic acid concentrations. Plasma concentrations are a more reliable marker of vitamin C status than dietary intake assessments. Urinary excretion of vitamin C occurs when the renal reabsorption threshold has been reached and plasma “saturation” occurs (between about 60 and 70μmol/L). Urinary excretion is influenced by recent dietary intake. Moreover, the sensitivity and specificity of this test is low, and 24h urine samples are generally required. Leukocyte ascorbic acid concentrations are considered a relatively reliable index of tissue ascorbic acid stores in healthy people; they are less responsive than plasma to short-term fluctuations in vitamin C intakes, but their assay is difficult and requires a larger blood sample. Furthermore, leukocyte ascorbic acid concentrations may not be a reliable marker of vitamin C status in some conditions, such as severe infection. The body pool size of vitamin C has been estimated using isotope dilution techniques. More research is needed to identify useful functional tests of vitamin C status.

17.1: Vitamin C (19.1)

17.2: Functions of Vitamin C (19.2)

17.3: Absorption and metabolism of vitamin C (19.3)

17.4: Deficiency of vitamin C in humans (19.4)

17.5: Food sources and dietary intakes of vitamin C (19.5)

17.6: Nutrient reference values for vitamin C (19.6)

17.7: Indices of vitamin C status (19.7)

17.8: Plasma ascorbic acid (19.8)

17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)

17.10: Ascorbic acid in erythrocytes and whole blood (19.10)

17.11: Urinary excretion of ascorbic acid and metabolites (19.11)

17.12: Ascorbic acid in other cells and fluids (19.12)

17.13: Body pool size (19.13)

17.14: Functional tests of vitamin C status (19.14)

This page titled 17: Vitamin C (Chapater 19) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.1: Vitamin C (19.1)

For centuries, the consumption of citrus fruit has been associated with the prevention of the vitamin C deficiency disease scurvy, but it was not until 1932 that the active compound was isolated (King & Waugh, 1932;Svirbely & Szent-Györgyi, 1932). A year later, ascorbic acid was characterized and became the first vitamin to be chemically synthesized. Humans are unable to synthesize vitamin C *in vivo* from glucose due to random genetic mutations, which occurred at an early stage in our evolution, and have resulted in loss of the terminal enzyme in the biosynthetic pathway, L-gulonolactone oxidase (Nishikimi & Yagi, 1991; Drouin et al., 2011).

The term “vitamin C” is used here to embrace both L-ascorbic acid and its oxidised form, L-dehydroascorbic acid; both have approximately equal vitamin C activity as dehydroascorbic acid can be readily reduced back to the biologically active ascorbic acid in the body. Ascorbic acid is the enolic form of an α-ketolactone (2,3-didehydro-threo-hexano-1,4-lactone), which in solution can be oxidized to the diketo form— dehydro-L-ascorbic acid; their structures are shown in Figure 19.1.

Under physiological conditions, vitamin C is present predominantly as the mono-anion, ascorbate.

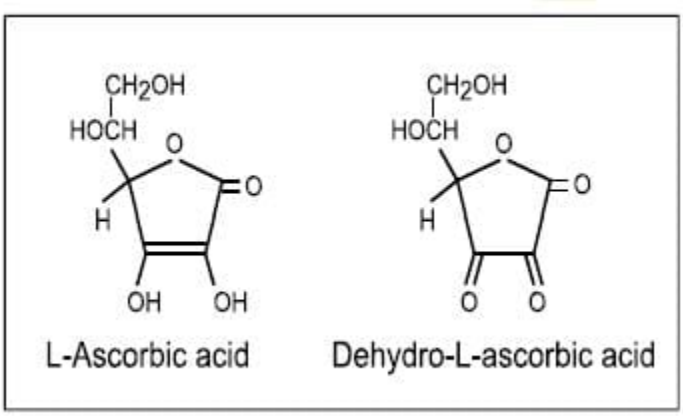


Figure 19.1. Structure of L-ascorbic acid and dehydroascorbic acid.

This page titled 17.1: Vitamin C (19.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.2: Functions of Vitamin C (19.2)

Ascorbic acid acts as an electron donating cofactor for a family of biosynthetic and regulatory metalloenzymes that carry out important hydroxylation reactions (Englard & Seifter, 1986; Kuiper & Vissers, 2014; Young et al., 2015). These include copper-containing mono-oxygenases and iron-containing dioxygenases. Of the latter, the most well-known is hydroxylation of proline and lysine residues during synthesis of collagen, the major connective tissue protein in the body. Ascorbic acid acts in these hydroxylation reactions by reducing the metal catalytic sites of procollagen proline 4-dioxygenase (EC 1.14.11.2) and procollagen lysine 5-dioxygenase (EC 1.14.11.4). Decreased activity of these two enzymes can lead to the defects in connective tissue that in turn cause the characteristic clinical features of scurvy. These include joint pain, tooth loss, bone and connective tissue disorders, and poor wound healing. Ascorbic acid is also involved in the hydroxylation of dopamine to noradrenaline, by acting as a cofactor for the enzyme dopamine-β-monooxygenase (EC 1.14.17.1). Noradrenaline is a neurotransmitter, and deficient dopamine hydroxylation may be associated with the mood changes, depression, and hypochondria that can occur with scurvy. The early symptoms of scurvy — fatigue, lethargy, and muscle weakness — are related to the role of ascorbic acid as a cofactor for maximal activity of two dioxygenases involved in carnitine biosynthesis (N-trimethyllysine dioxygenase; EC 1.14.11.8, and γ-butyrobetaine dioxygenase; EC 1.14.11.1). Carnitine is essential for the transport of activated long-chain fatty acids into the mitochondria, where they are converted to energy by way of β-oxidation (Englard & Seifter, 1986).

Several other enzymes depend on ascorbic acid as a cofactor, although their specific role in the development of scurvy is still unclear. These include peptidylglycine α-amidating monooxygenase (EC 1.14.17.3), an enzyme required for the synthesis of numerous amidated peptide hormones which have multiple functions throughout the body (Englard & Seifter, 1986). More recent discoveries have included the hydroxylases that regulate the concentration and activity of the transcription factor hypoxia-inducible factor-1α (i.e. hypoxia-inducible factor — proline dioxygenase; EC 1.14.11.29, and hypoxia-inducible factor — asparagine dioxygenase; EC 1.14.11.30) (Kuiper & Vissers, 2014), and the ten-eleven translocation (TET) hydroxylases and histone demethylases that modify epigenetic marks on DNA and histones (Young et al., 2015). Together, these epigenetic and transcription regulating enzymes have the potential to up- and down-regulate thousands of genes in the body, likely contributing to vitamin C's pleiotropic functions in human health and disease.

Ascorbic acid also has several nonenzymatic functions based on its role as a reducing agent and antioxidant (i.e. electron donation). For example, ascorbic acid enhances the gastrointestinal absorption of nonheme iron, through its ability to reduce intraluminal iron to its more absorbable ferrous form (Hallberg et al., 1987). Ascorbic acid is a potent antioxidant, with the ability to quench a wide range of reactive oxygen species, including neutralising phagocyte-derived oxidants, thus attenuating tissue damage (Carr & Frei, 1999a). Ascorbic acid provides important antioxidant protection to plasma lipids and lipid membranes through direct scavenging of reactive oxygen species and through the regeneration of vitamin E from its oxidised form (α-tocopheroxyl radical). Ascorbic acid can also regenerate the enzyme cofactor tetrahydrobiopterin from its oxidised form (dihydrobiopterin), thus potentially aiding synthesis of dopamine, serotonin and nitric oxide. At high intakes, vitamin C has been associated with protection against cardiovascular disease, cataracts, and cancer (Carr & Frei, 1999b), and a variety of immune-related functions (Carr & Maggini, 2017). Additional research, however, is needed to establish causal relationships between vitamin C and these degenerative diseases and immunocompetence.

This page titled 17.2: Functions of Vitamin C (19.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.3: Absorption and metabolism of vitamin C (19.3)

Vitamin C is absorbed in the small intestine by an active sodium-dependent vitamin C transporter (SVCT-1) that is dose dependent and can become saturated (Savini et al., 2008). About 70%–90% of vitamin C is absorbed when daily intakes range from 30 to 180mg, but absorption falls to ≤ 50% at intakes above 1g/d (Kallner et al., 1979; Levine et al., 1996). Indeed, the osmotic diarrhea and intestinal discomfort sometimes experienced by persons ingesting large gram doses of vitamin C is due to the presence of a large amount of unabsorbed vitamin C. Regulation of the whole body vitamin C content is achieved in part by this dose-dependent intestinal absorption of vitamin C. Renal action to conserve vitamin C (via SVCT-1) or excrete unmetabolized vitamin C also plays an important part. Tissue uptake occurs primarily by the SVCT-2 isoform (Savini et al., 2008). Different tissues uptake and retain vitamin C to variable extents, with neuronal and glandular tissues containing the highest concentrations (Hornig, 1975), indicating vital roles for vitamin C in these tissues. Catabolism of ascorbic acid in humans occurs through oxidation to dehydroascorbic acid and, in the absence of sufficient reducing equivalents, subsequent hydrolysis to diketogulonic acid. This can then be oxidised to smaller molecules, such as oxalate, that are excreted in urine. However, it should be noted that dehydroascorbic acid is readily taken up via cell membrane glucose transporters (GLUTs) and rapidly reduced back to ascorbic acid via various chemical and enzymatic pathways (Rumsey et al., 1997; Washko et al., 1993), thus limiting dehydroascorbic acid concentrations in vivo. With large intakes, most of the vitamin C is excreted in its unmetabolized form (Levine et al., 1996).

This page titled 17.3: Absorption and metabolism of vitamin C (19.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.4: Deficiency of vitamin C in humans (19.4)

Clinical manifestations of scurvy, such as weakness, petechial hemorrhages, gingival bleeding, ecchymoses and subperiosteal hemorrhage, anemia, and defects in bone development in children, are relatively rare in high-income countries. Nevertheless, vitamin C deficiency has been observed in institutionalized elderly subjects, likely arising from inadequate intake of vitamin C (Carr & Rowe, 2020). Alcoholic patients, those on restrictive diets and people in refugee camps are also vulnerable to vitamin C deficiency. Infantile scurvy is rare because breast milk provides adequate amounts of ascorbic acid and infant formulas are fortified with sufficient ascorbic acid to prevent the disease.

Subclinical vitamin C deficiency may develop secondarily to some disease states such as severe respiratory infections, cancer, gastrointestinal disorders, cardiometabolic disorders and inflammatory diseases. This is likely partly due to the enhanced oxidative stress and inflammation associated with many of these diseases, as well as the use of certain drugs (e.g. aspirin) or chemotherapeutic agents that appear to affect the bioavailability or metabolism of ascorbic acid (Basu, 1982; Carr & Cook, 2018).

The functional health consequences of both mild and moderate vitamin C deficiency are not well characterized. Gingival inflammation, fatigue, and irritability have been reported in moderate vitamin C deficiency induced experimentally (Leggot et al., 1986; Levine et al., 1996).

This page titled 17.4: Deficiency of vitamin C in humans (19.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.5: Food sources and dietary intakes of vitamin C (19.5)

Vitamin C is synthesised in plants. However, the vitamin C content of different foods is highly variable (Carr & Rowe, 2020). Values can also depend on the growing conditions, season of the year, stage of maturity, location, and storage time prior to consumption. Exposure to high temperatures, oxidation, or cooking in large amounts of water all reduce the vitamin C content of foods.

Major food sources of vitamin C in the diets of most high-income countries are fresh fruits and fruit juices and some fresh vegetables; they can contribute up to 90% of the vitamin C intake. Citrus and kiwifruit are particularly rich sources of vitamin C. Of the vegetables, potatoes can be one of the most important sources of vitamin C in the diets of adults, for example in the U.K. (Gregory et al.,1990).

Most staple foods, e.g. rice, wheat, maize and other grains, contain negligible amounts of vitamin C and meat, fish, eggs, and dairy products are also poor sources (Carr & Rowe, 2020). In low-income countries, the supply of food sources of vitamin C is often seasonal, so intakes may vary widely. In The Gambia, for example, intakes ranged from zero to 115mg/d in one community, depending on the seasonal availability of mangos, oranges, and grapefruit (Bates et al., 1994).

In humans, the bioavailability of vitamin C from food sources is comparable to that from supplements and is not affected by the type of food consumed(Carr & Vissers, 2013).

This page titled 17.5: Food sources and dietary intakes of vitamin C (19.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.6: Nutrient reference values for vitamin C (19.6)

Several expert groups including the Institute of Medicine (IOM) in the U.S. and the European Food Safety Authority (EFSA) have derived values for Average Requirements (ARs) and Recommended Intakes (RIs) for vitamin C. The AR (U.S. equivalent, Estimated Average Requirement) is set for a particular group of individuals and is used to assess the prevalence of adequate intakes of vitamin C within that group. The RIs (U.S. equivalent, Recommended Daily Allowance (RDA); EFSA equivalent, Population Reference Intake) are derived from the ARs and its distribution and are used to assess intakes and plan diets for individuals.

Recommendations for vitamin C intakes vary greatly around the world, with recommended intakes (RIs) for individuals ranging from as low as 40mg/d in the U.K. to 110mg/d by EFSA (2017). In the U.S., the RDAs established by the IOM in 2000 are 75mg/d and 90mg/d for women and men, respectively (Levine et al., 1996; IOM, 2000). The large discrepancies in the RIs have resulted in calls for harmonization (Carr & Lykkesfeldt, 2020). Higher vitamin C intakes are recommended by all expert groups for pregnant and lactating women due to the needs of the developing fetus and the transfer of vitamin C to breast milk. Furthermore, some countries (e.g., the U.S.) recommend higher intakes for smokers due an enhanced oxidative burden and resultant increase in turnover of vitamin C. Consequently, RIs for smokers may be increased by +20mg/d to +80mg/d, depending on the country.

Upper levels of intake (ULs) have also been derived for vitamin C by some expert groups. The ULs represent daily intakes that if consumed chronically over time will have a very low risk of causing adverse effects. Intakes from all sources of vitamin C are considered: food and nutrient supplements. Because vitamin C is water soluble, however, any excess not required by the body is readily excreted, therefore, there is no known upper toxic concentration. Nevertheless, some countries have set ULs ranging from 1–2g/d for adults based on potential gastrointestinal disturbances at very large (> 4g/d) doses. For more details on the application of these nutrient reference values, see Chapter 8b.

In addition to diarrhea and other gastrointestinal disturbances, possible adverse effects of excessive intakes of vitamin C include an increase in the absorption of dietary iron potentially leading to iron overload in people with hemochromatosis, and hemolytic anaemia in people with glucose-6-phosphate dehydrogenase deficiency due to an inability to attenuate hydrogen peroxide generation. Current evidence is insufficient to implicate high intakes of vitamin C as a risk factor in oxalate stone formation. Nevertheless, individuals with hemochromatosis, glucose-6-phosphate dehydrogenase deficiency, and renal disorders, may be susceptible to adverse effects from excessive intakes of vitamin C (IOM, 2000).

This page titled 17.6: Nutrient reference values for vitamin C (19.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.7: Indices of vitamin C status (19.7)

There are no reliable functional tests of vitamin C status. Instead, static biochemical tests, particularly the measurement of plasma or leukocyte ascorbic acid concentrations, are most frequently used to assess vitamin C status. These static biochemical tests are discussed in detail below.

This page titled 17.7: Indices of vitamin C status (19.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.8: Plasma ascorbic acid (19.8)

Ascorbic acid is transported in the plasma but is not bound to any protein; almost all is present as the ascorbate monoanion. Plasma (or serum) ascorbic acid concentrations are the most frequently used and practical index of vitamin C status in studies of individuals and populations. Concentrations are influenced by recent intake of the vitamin, making fasting blood samples essential.

19.8.1 Ascorbic acid pharmacokinetics

Plasma ascorbic acid concentrations exhibit a characteristic sigmoidal relationship with intake, as shown in Figure 19.2. In adults, the steepest change is evident at intakes between about 30 and 90mg/d of vitamin C. When intakes are below about 20mg/d, most of the vitamin C enters the tissues, so there is very little available for circulation in the blood. For intakes greater than 100mg/d, plasma concentrations tend to plateau between 60 and 70µmol/L, concentrations representing the renal threshold and plasma saturation (Figure 19.2). When intakes exceed about 200mg/d, intestinal absorption of vitamin C becomes the limiting factor (Levine et al., 1996).

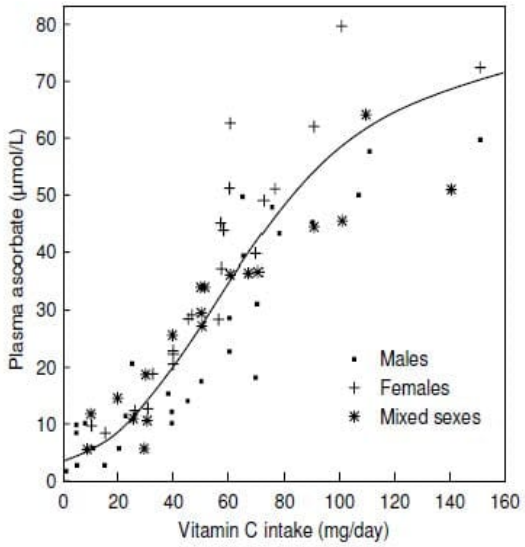


Figure 19.2. The relationship of plasma ascorbate relative to daily vitamin C intake. Redrawn from Bates et al. (1997).

As a result of this threshold effect with only limited intestinal absorption at intakes > 200mg/d, and the excess circulating ascorbic acid being excreted in the urine, plasma ascorbic acid concentrations cannot identify people who are regularly consuming excessive amounts of vitamin C. Hence, it is not surprising that in the U.S. National Health and Nutrition Examination Survey (NHANES II), a correlation between plasma ascorbic acid and vitamin C intakes was only reported for people not taking vitamin C supplements (Loria et al., 1998). Vitamin C is one of the best markers for fruit and vegetable intake, particularly fruit intake (Block et al., 2001; Drewnowski et al., 1997). However, strong correlations between dietary intakes of vitamin C and plasma ascorbic acid concentrations have only been reported when habitual dietary intakes of ascorbic acid are relatively modest (30–80mg/d) (Bates et al., 1979).

People consuming chronically low intakes of vitamin C likely have plasma ascorbic acid concentrations that reflect the ascorbic acid content of the body (Jacob et al., 1987). Indeed, in such circumstances, plasma concentrations are probably as accurate an indicator as concentrations in leukocytes. However, the relationship may be obscured in epidemiological studies by other factors that affect the absorption or metabolic turnover of vitamin C. In NHANES III, for example, only 50% of adult men had serum ascorbic acid concentrations > 38µmol/L, although more than 75% had dietary intakes higher than the estimated average requirement (i.e., 75mg/d) (IOM, 2000). Such a discrepancy may be due, in part, to the high proportion of U.S. adult males who smoke or who are exposed to environmental tobacco smoke.

19.8.2 Factors affecting plasma ascorbic acid

Cigarette smoking has a marked effect on plasma ascorbic acid concentrations. Many studies report lower plasma ascorbic acid concentrations in smokers (Carr & Rowe, 2020), as well as in those regularly exposed to environmental cigarette smoke (Figure 19.3). The mechanism controlling the influence of cigarette smoking on plasma ascorbic acid concentrations appears to be related to the higher metabolic turnover of vitamin C in smokers than in nonsmokers (Kallner et al., 1981), which is probably induced by increased oxidative stress from substances in smoke. Certainly, studies using isotopically labeled ascorbic acid show an increased metabolic turnover of ascorbic acid in smokers (70mg/d versus 36mg/d for nonsmokers), resulting in a 40% greater requirement for vitamin C (IOM, 2000). Smokers also tend to have lower dietary intake of vitamin C, further contributing to a poor vitamin C status (Schectman et al., 1989).

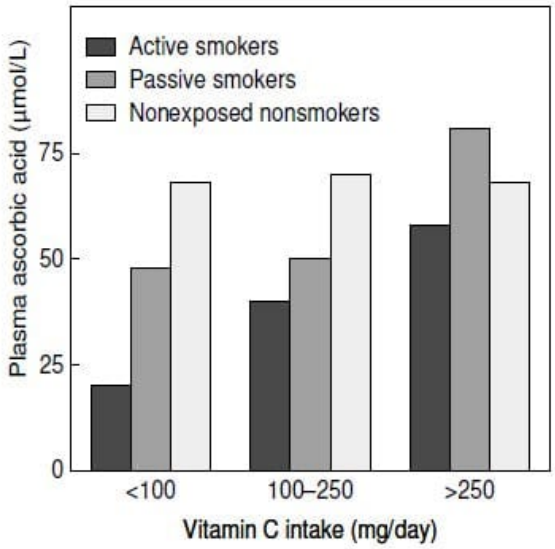


Figure 19.3. The effects of cigarette-smoke exposure on plasma ascorbic acid concentrations in 141 premenopausal women who were active or passive smokers, or nonexposed nonsmokers. Three concentrations of dietary vitamin C intake are represented in the results. Data from Tribble et al. (1993).

Sex influences plasma ascorbic acid concentrations; women have higher plasma ascorbic acid concentrations than men who consume similar intakes of vitamin C. However, this is likely partly explained by volumetric dilution due to differences in fat-free mass (Jungert & Neuhauser-Berthold, 2015), a premise supported by a lack of sex differences in plasma ascorbic acid concentrations before adolescence. For example, in NHANES III, higher serum ascorbate concentrations in women than in men were only apparent in those aged > 19y; no sex-related differences were reported in younger age groups (IOM, 2000). Thus, sex differences in lean body mass are likely to be a contributing factor (Blanchard, 1991).

Age-differences in plasma ascorbic acid concentrations have been reported. The elderly, especially those who are institutionalized or housebound, may have lower concentrations than those of younger persons (Carr & Rowe, 2020). This trend has been attributed in part to lower dietary intakes because of problems of poor dentition or chronic diseases that may influence absorption or metabolism of the vitamin. Nevertheless, at present, there is no evidence that the elderly have higher vitamin C requirements than those for young adults. Indeed, NHANES IV data has indicated a U-shaped curve for circulating vitamin C concentrations by age (Schleicher et al., 2009).

Pregnancy lowers plasma ascorbic acid concentrations due to increasing body weight and hemodilution, as well as active transfer to the fetus, especially during the last trimester. Lactation results in even higher requirements due to active transfer of an average of 40mg/d vitamin C through breast milk to the growing infant (IOM, 2000).

Higher body weight is typically associated with lower vitamin C status in epidemiological studies. This is likely due to volumetric dilution, whereby an identical vitamin C intake will result in lower vitamin C concentrations in a person of higher bodyweight relative to a person of lower bodyweight (Block et al., 1999). People with central obesity are likely to have even higher requirements due to enhanced oxidative stress and inflammation (Carr et al., 2022). Despite the obesity pandemic in many regions of

the world, and the use of bodyweight to estimate vitamin C recommendations for women and children, there are as yet no vitamin C recommendation categories for large or overweight individuals.

Acute or chronic infection lowers circulating ascorbic acid concentrations. The more severe the infection, the lower the vitamin C status, as evidenced by the high prevalence of hypovitaminosis C and deficiency in hospitalised patients with pneumonia and sepsis (Carr, 2020).

19.8.3 Interpretive criteria for plasma ascorbic acid

Protracted intakes of vitamin C < 20mg/d cause plasma ascorbic acid concentrations to decline rapidly to deficient concentrations (i.e. 11µmol/L or less) (Hodges et al., 1971; Jacob et al., 1987). At the lower plasma ascorbic acid concentrations, clinical signs of scurvy such as follicular hyperkeratosis, swollen or bleeding gums, petechial hemorrhages, and joint pain occur (Hodges et al., 1969). Therefore, plasma ascorbic acid concentrations ≤ 11µmol/L are taken as indicative of vitamin C deficiency.

The cut-off values for plasma ascorbic acid concentrations used to distinguish different categories of deficiency are poorly defined. NHANES II defined "clinically based categories" of plasma ascorbate concentrations in terms of mg/dL, i.e. deficient < 0.2, low 0.2–0.39, normal 0.4–0.99, and saturated ≥ 1.0mg/dL (equating to < 11, 11–22, 23–56, and ≥ 57µmol/L, respectively) (Loria et al., 1998). Of note, the low (hypovitaminosis C) cut-off is when the "sub-clinical" signs and symptoms of vitamin C deficiency start to be observed, e.g. changes in mood and energy levels.

More recently, the European Food Safety Authority (EFSA) has defined ≥ 50µmol/L as "adequate" (Table 19.1; EFSA, 2013). Well-conducted pharmacokinetic studies in men and women suggest 70µmol/L as a cut-off for "saturation", although there is variation between individuals (Levine et al., 1996; Levine et al., 2001). The criteria in Table 19.1 have been used in more recent NHANES analyses (Crook et al., 2021), that indicated that smoking, male sex and higher body weight/BMI were common indicators for lower vitamin C status.

Table 19.1. Current interpretive criteria for plasma ascorbic acid concentrations.

Status	Plasma Ascorbic Acid (µmol/L)
Deficient	< 11
Hypovitaminosis C	< 23
Inadequate	< 50
Adequate	≥ 50
Saturating	≥ 70

Of note, the mean vitamin C status of many high-income populations globally is typically around 50µmol/L, with the noted sex differences (Rowe & Carr, 2020). The prevalence of hypovitaminosis C and outright deficiency also tends to be relatively low in these countries. In contrast, many low-middle income populations typically have lower mean vitamin C status, and a higher prevalence of hypovitaminosis C and deficiency, largely explained by lower dietary intakes. Vitamin C status and intake data from national surveys (e.g. U.S. NHANES and U.K. National Diet and Nutrition Survey) can be found in the review of Rowe and Carr (2020) and the cited references.

19.8.4 Measurement of plasma ascorbic acid

Fasting blood samples are required for plasma ascorbic acid analysis. The samples need to be preserved immediately after collection to avoid degradation of the ascorbic acid. Metaphosphoric acid (or trichloroacetic acid) is typically used to precipitate protein and to stabilize the ascorbic acid in the samples prior to analysis. Additionally, a reducing agent, such as dithiothreitol (DTT), can be added to preserve the ascorbic acid in a reduced state. The reducing agent TCEP (tris(2-carboxyethyl)phosphine) is versatile in that it can also work in acidified solutions with longer incubation times. A metal chelator, such as DTPA (diethylenetriaminepentaacetic acid), can be added to attenuate metal ion-dependent oxidation of ascorbic acid (particularly important if samples are hemolysed). It should be noted that EDTA (ethylenediaminetetraacetic acid), although a commonly used metal-chelator, is redox-active and can catalyse the oxidation of ascorbic acid if the samples are not handled appropriately (Pullar et al., 2018).

Plasma samples for ascorbic acid analysis that have been prepared appropriately and frozen at −20°C are stable for several

weeks and at –70°C for at least 1y (Margolis & Duewer, 1996). If the samples cannot be processed immediately, whole blood can be stored for up to 8h at 4°C before processing (Galan et al., 1988). Care needs to be taken with choice of anti-coagulants for collection of blood samples as the anti-coagulant EDTA, although a metal-chelator, remains redox active and can result in loss of ascorbic acid if the blood or plasma samples are not kept at 4°C (Pullar et al., 2018).

Several methods are available for measuring vitamin C in both the reduced form or as total ascorbic acid. The older colourimetric or fluorometric assays use predominantly 2,4-dinitrophenylhydrazine or o-phenylenediamine. These methods, however, have a number of limitations (Washko et al., 1992). Their sensitivity and specificity is low, and they are often subject to interference by other biological substances and yield falsely high readings at low ascorbic acid concentrations. In some circumstances, inadvertent oxidation of ascorbic acid or hydrolysis of dehydroascorbic acid may occur.

The method of choice for measuring ascorbic acid in plasma samples is high-performance liquid chromatography (HPLC); only 50µL of plasma is required for the measurement. Of the HPLC methods available, electrochemical detection is preferred, with its high selectivity and sensitivity. However, this method does require a dedicated instrument and an experienced operator (Washko et al., 1989). Other simpler methods use HPLC coupled with an ultraviolet detector, but these methods can have a lower sensitivity and specificity (Washko et al., 1992).

Analysis of dehydroascorbic acid by HPLC is more difficult because direct electrochemical or UV detection of dehydroascorbic acid is not possible. Instead, samples must first be analysed for ascorbic acid and then reduced for measurement of ascorbate plus dehydroascorbic acid. Dehydroascorbic acid is then determined by subtraction. *In vivo* concentrations of dehydroascorbic acid are generally very low (e.g. < 2µmol/L) due to rapid uptake (via GLUTs) and intracellular reduction to ascorbic acid by cells of the vasculature. Therefore, reports of higher plasma (or serum) concentrations of dehydroascorbic acid in different disease states are likely due to artefactual *ex vivo* oxidation often associated with use of the older spectrophotometric methods and/or inappropriate handling and processing of the samples prior to analysis (Pullar et al., 2018).

This page titled 17.8: Plasma ascorbic acid (19.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)

Leucocytes include lymphocytes, monocytes, and three classes of granulocytes: polymorphonuclear leukocytes or neutrophils, eosinophils, and basophils. These cell types differ in their concentration of ascorbic acid and possibly in their response to supplemental vitamin C (Levine et al., 1996; Levine et al., 2001), thus complicating the interpretation of ascorbic acid concentrations in leukocytes.

19.9.1 Leukocyte ascorbic acid

Concentrations of ascorbic acid in plasma and leukocytes correlate relatively well (Figure 19.4; Sauberlich et al., 1989), even though ascorbic acid concentrations in leukocytes are at least 14 times greater than those in plasma (Levine et al., 1996; Levine et al., 2001). The ascorbic acid concentrations of mixed leukocytes range from 90–300nmol/10^8 cells in adults (Omaye et al., 1987), and depend in part on the heterogeneous nature of the leukocytes, and the methods used to isolate and analyse them.

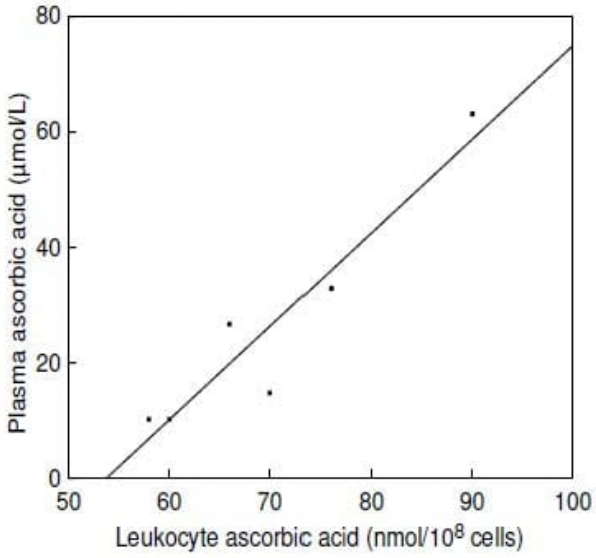


Figure 19.4. The relationship of mean plasma and leukocyte ascorbic acid concentrations measured at various stages during a depletion and repletion study of a group of six adult females. Data from Sauberlich et al., 1989),

Leukocyte ascorbic acid concentrations are commonly believed to be a more reliable index of tissue stores of ascorbic acid than are the corresponding concentrations in the plasma, erythrocytes, or whole blood (Turnbull et al., 1981). Leukocyte ascorbic acid concentrations are less responsive than plasma to short-term fluctuations in recent vitamin C intakes. In the study shown in Figure 19.5, leukocyte ascorbic acid concentrations decreased about 33%, compared with the much larger and more rapid decline in plasma. Similar trends have been reported by others (Jacob et al., 1987). Thus, they provide a less sensitive measure of lower intakes.

Concentrations of ascorbic acid in leukocytes reflect changes in tissue ascorbic acid concentrations and the ascorbic acid body pool, but not dietary intakes. Animal studies (i.e., monkeys and guinea pigs) have confirmed that leukocyte ascorbic acid concentrations provide an accurate reflection of ascorbic acid concentrations in the liver and body pool (Omaye et al., 1979).

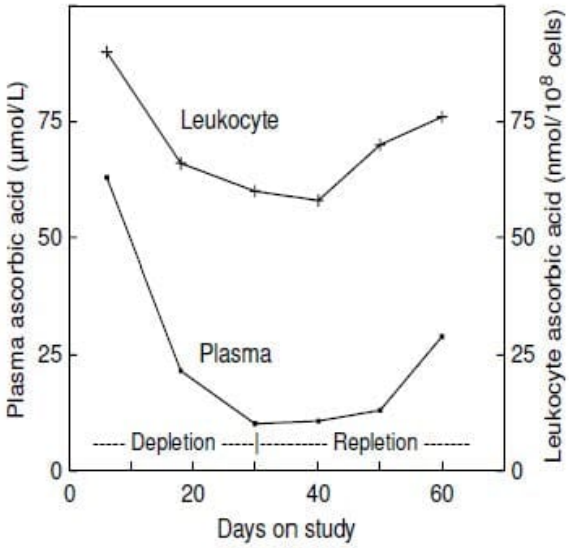


Figure 19.5. Mean plasma and leukocyte ascorbic acid concentrations during a depletion and repletion study of a group of six adult females. Ascorbic acid intakes were ad libitum for days 1–6; 0mg/d for days 6–30; 30mg/d for days 31–40; 60mg/d for days 41–50; and 90 mg/d for days 51–60. Data from Sauberlich et al., 1989),

Several factors can influence the concentrations of ascorbic acid in leukocytes. It has been proposed that the platelet/leukocyte ratio can influence leukocyte ascorbic acid values (Vallance, 1986). Platelets have comparable ascorbic acid concentrations to mononuclear cells (Levine et al., 2001), and should ideally be separated from leukocytes prior to leukocyte ascorbic acid assays (Vallance, 1986). Choice of the anticoagulant used may influence binding of platelets to leukocytes (Healy & Egan, 1984).

Some additional non-dietary factors that influence concentrations of ascorbic acid in leukocytes are smoking, sex, infection, and certain drugs, as discussed for plasma ascorbic acid concentrations. In some cases, these effects are due to alterations in cell populations (e.g., acute infection) or to differences in ascorbic acid uptake (Lee et al., 1988).

Finally, concentrations of vitamin C in phagocytic cells, such as neutrophils, can be influenced by activation of their oxidative burst. This results in the generation of reactive oxygen species which oxidise extracellular ascorbic acid to dehydroascorbic acid that is readily transported into the cells via membrane GLUTs, where it is then rapidly reduced back to ascorbic acid (Washko et al., 1993). Higher than anticipated concentrations of vitamin C have been observed in leukocytes isolated from patients with severe sepsis, which is a condition characterised by systemic oxidative stress and activation of neutrophils (Carr et al., 2021).

19.9.2 Ascorbic acid in specific cell subsets

The specific cell type that is most useful for assessing vitamin C status is presently uncertain. As noted earlier, both ascorbic acid concentrations in individual leukocytes and their response to vitamin C supplementation vary. Mononuclear cells (lymphocytes and monocytes) have concentrations that are two- to threefold higher than granulocytes (e.g., neutrophils). Furthermore, the ascorbic acid pool in mononuclear cells is depleted more slowly than that from other blood compartments (Jacob, 1990). In addition, it can be difficult to obtain homogeneous and reproducible fractions of specific cell types, so that their reported ascorbic acid concentrations can vary widely.

Lymphocyte ascorbic acid concentrations of 120–250nmol/10^8 cells have been reported for healthy adults not consuming supplemental vitamin C(Evans et al., 1982). Jacob et al. (1991) measured lymphocyte ascorbic acid concentrations in a depletion-repletion study of healthy men that was designed to induce moderate vitamin C deficiency. The mean baseline lymphocyte ascorbic acid concentration in the men was 209nmol/10^8 cells. After 60d of depletion, lymphocyte ascorbic acid concentrations fell significantly and consistently to a mean concentration of 87nmol/10^8 cells. Moreover, the concentrations distinguished between the group of participants receiving 5, 10, or 20mg ascorbic acid (74–145 nmol/10^8 cells) versus those participants receiving 60 or 250mg/d (182–261nmol/10^8 cells). Strong correlations of plasma and lymphocyte ascorbic acid concentrations were noted within the individual subjects when ascorbic acid intakes ranged from 5–250mg/d (Table 19.2) There was no evidence

of scorbutic symptoms in the depletion phase of this study, but there was evidence of greater oxidative damage, based on alterations in indices of oxidant status (Jacob et al., 1991).

Table 19.2. Ascorbic acid concentrations of healthy men after periods of various ascorbic acid intakes. Ascorbic acid concentrations not sharing the same subscript letter are significantly different (< 0.05) by t-test. † Participants consumed a supplement of 250mg a deleted because of platelet contamination. Data From Jacob et al., (1991)

Days	Ascorbic acid intake (mg/d)	n	Plasma (µmol/L)
4	250†	8	59.6 ± 3.3[a]
32	5	8	6.6 ± 0.3[b]
28	10	4	5.3 ± 0.3[b]
28	20	4	5.9 ± 0.5[b]
28	60	3	26.7 ± 8.8[c]
28	250	5	64.2 ± 4.5[a]

In later ascorbic acid depletion-repletion studies in men and women, Levine et al. (1996, 2001) measured concentrations of ascorbic acid in purified neutrophils, monocytes and lymphocytes. Neutrophil ascorbic acid concentrations increased markedly in response to supplemental vitamin C intakes of between 30 and 100mg/d but showed little further increase at higher doses. This trend is similar to that observed for plasma ascorbic acid concentrations (Figure 19.2), although the latter reach a plateau at higher daily doses (about 200mg/d). The concentrations of ascorbic acid in purified monocytes and lymphocytes followed a similar trend, again reaching a plateau at doses of 100mg/d.

19.9.3 Interpretive criteria for leukocytes

Confusion exists over the interpretive criteria used for leukocyte ascorbic acid concentrations, in part because they can be expressed in different ways: per 10^8 cells, per mL, per DNA concentration, or per unit of protein. This makes comparisons among laboratories difficult. The most common method is in terms of cell numbers (nmol/10^8cells). The heterogeneous nature of the leukocytes and various technical difficulties with their analyses, are further complicating factors.

Table 19.3 presents the interpretive criteria used by Jacob (1994) and given in Sauberlich (1999), expressed in terms of cell numbers. When expressed in this way, clinical signs of scurvy, such as swollen or bleeding gums or petechial hemorrhages, have been associated with leukocyte concentrations of about 11nmol/10^8 cells. Even at concentrations as high as 50nmol/10^8 cells, scorbutic changes including inflammation, tenderness, bleeding of the gums, and petechiae have sometimes been described (Sauberlich et al., 1989). For this reason, the cut-off shown in Table 19.3, indicative of deficiency, is < 57nmol/10^8 cells.

Table19.3. Interpretive criteria for leukocyte ascorbic acid concentrations. From Jacob (1994).

Status	Mixed leukocytes (nmol/10^8 cells)	Mononuclear leukocytes
Deficient	< 57	< 114
Low	57–114	114–142
Adequate	> 114	> 142

Cut-off values for ascorbic acid concentrations in specific cell types such as neutrophils, monocytes, and lymphocytes are less certain. Levine et al. (1996, 2001) showed that intra-cellular ascorbic acid concentrations in neutrophils were less than half of corresponding mononuclear leukocyte (lymphocyte and monocyte) concentrations. The interpretation of leukocyte ascorbic acid concentrations may be particularly difficult in surgical patients or in cases of severe infection; these conditions are often associated with leukocytosis, activation of neutrophils, and systemic oxidative stress (Schorah et al., 1986; Vallance, 1988; Carr et al., 2021).

19.9.4 Measurement of leukocyte ascorbic acid

The isolation and assay of leukocytes for ascorbic acid is technically difficult, and presently requires relatively large blood samples (2–5mL), making the assay unsuitable for serial measurements on infants. Leukocytes can be isolated with density-gradient sedimentation (Boyum, 1968). Monoclonal antibodies to various cell types can also be used (Field, 1996). Analysis of ascorbic acid in leukocytes or specific cell types is best performed by HPLC with electrochemical detection, after deproteinization (Washko et al., 1989).

This page titled 17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.10: Ascorbic acid in erythrocytes and whole blood (19.10)

Erythrocyte ascorbic acid concentrations are not widely used as an index of ascorbic acid status. Concentrations are roughly comparable to plasma but only respond to changes in vitamin C intake over a narrow range. Hence, they are not as sensitive as plasma ascorbic acid concentrations (Hodges et al., 1971). Because erythrocyte concentrations are less responsive to recent dietary intake, they could potentially be used to assess ascorbic acid status in non-fasting individuals (Pullar et al., 2020). However, both within- and between-variation and analytical variance can be greater than for plasma ascorbic acid (Jacob et al., 1987). Erythrocyte ascorbic acid is technically more challenging to measure due to potential for oxidation of ascorbic acid by hemoglobin-associated iron; the hemoglobin is ideally removed via centrifugal filter units (Li et al., 2012). It should also be noted that conditions of vitamin C deficiency can result in enhanced erythrocyte fragility and hemolysis (Tu et al., 2015).

Whole blood ascorbic acid concentrations have also been investigated in ascorbic acid depletion-repletion studies. Again, concentrations in whole blood are not as sensitive as plasma for assessing ascorbic acid depletion (Jacob et al., 1987). Concentrations < 17µmol/L are considered deficient because they have been associated with scorbutic signs (Sauberlich et al., 1989); those of 17–28µmol/L are considered low. Concentrations > 28µmol/L are interpreted as acceptable (Sauberlich, 1999). Presently, whole blood ascorbic acid concentrations are rarely used.

This page titled 17.10: Ascorbic acid in erythrocytes and whole blood (19.10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.11: Urinary excretion of ascorbic acid and metabolites (19.11)

Urine is the major excretory route for absorbed ascorbic acid. Most ascorbic acid is excreted in its unmetabolized form (Levine et al., 1996). When intakes are > 1g/d, there is some increase in oxalate excretion (Olson & Hodges, 1987). However, reports of high oxalate excretion are likely due to artefactual *ex vivo* oxidation of excreted ascorbic acid prior to analysis (Chalmers et al., 1985), therefore careful handling of urine samples is required.

Urinary excretion of ascorbic acid reflects recent dietary intake. Experimental depletion-repletion studies have demonstrated that concentrations in the urine decline progressively with increasing depletion of vitamin C until, in persons with scurvy, concentrations are undetectable (Jacob et al., 1987). Nevertheless, urinary excretion is not a very sensitive index of ascorbic acid status; differences between persons with adequate or deficient intakes of ascorbic acid are small. For example, in the depletion-repletion study of 11 young men shown in Figure 19.6, although urinary ascorbic acid excretion was significantly lower during the depletion periods (weeks 3–6) than in the initial 2wk 65mg/d baseline period, the differences were small.

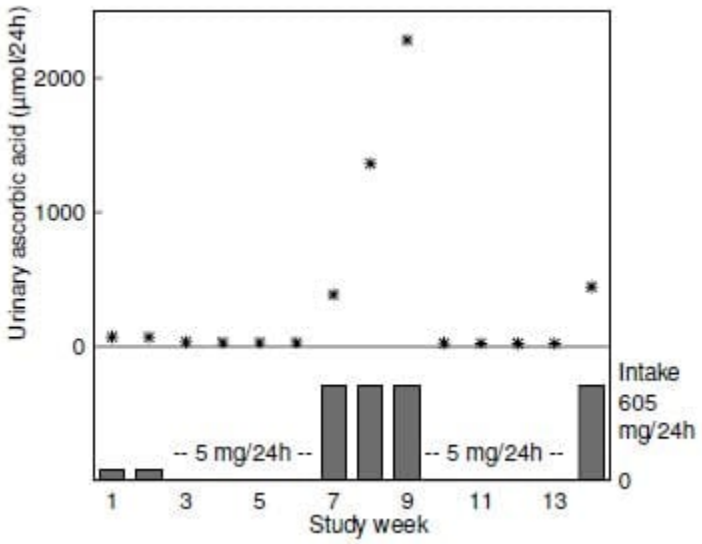


Figure 19.6. Mean urinary ascorbic acid during times of varied ascorbic acid intake. Data from a study of 11 adult men by Jacob et al. (1987).

More recent pharmacokinetic investigation has indicated a renal threshold of about 60µmol/L plasma ascorbic acid for a majority of healthy individuals (Ebenuwa et al., 2022), a concentration attained by intakes of ≥ 100mg/d (Levine et al., 1996, 2001). Higher concentrations of ascorbic acid are observed in urine when plasma "saturation" is attained (i.e. intakes of ≥ 200mg/d). The vitamin C requirements of an individual can potentially be determined by administering a test dose of vitamin C and monitoring the resultant excretion.

The specificity of urinary ascorbic acid, however, is also low. The drugs aminopyrine, aspirin, barbiturates, hydantoins, and paraldehyde can all increase urinary ascorbic acid excretion (Sauberlich, 1981). An additional disadvantage of this test as a measure of ascorbic acid status in humans is the requirement for 24h urine specimens. The latter, although useful in clinical or research settings, are less practical in field studies.

19.11.1. Measurement of urinary ascorbic acid

Ascorbic acid is unstable in urine and samples should ideally be kept at 4°C and stabilized with a metal chelator (e.g. DTPA) and acidification (e.g. with metaphosphoric acid) as soon as possible after collection. HPLC with electrochemical detection is

the preferred method of analysis. Values can be presented as amount excreted per hour or day, or relative to urine creatinine concentrations for spot collections, to account for urine concentration.

This page titled 17.11: Urinary excretion of ascorbic acid and metabolites (19.11) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.12: Ascorbic acid in other cells and fluids (19.12)

Ascorbic acid has been detected in a number of other body fluids. The concentrations in saliva and alveolar lining fluid are generally relatively low and do not appear to correlate with concentrations of ascorbic acid in plasma, leukocytes, or dietary intakes (Leggott et al., 1986; Jacob et al., 1987; Bui et al., 1992). In contrast, ascorbic acid concentrations in cerebrospinal fluid are typically higher than those found in plasma (Tallaksen et al., 1992). Ascorbic acid is able to bypass the blood-brain barrier via SVCTs in the choroid plexus (Angelow et al., 2003), and is believed to have important functions in the central nervous system (May, 2012). Of note, concentrations of ascorbic acid in seminal fluid can be up to 10-fold higher than those in plasma (Thiele et al., 1995). Lower seminal ascorbic acid concentrations were observed in infertile men and this appeared to be related to elevated oxidative stress (Thiele et al., 1995; Lewis et al., 1997).

Buccal cell ascorbic acid concentrations have been investigated as a potential biomarker of vitamin C status. In a vitamin C depletion-repletion study in healthy men, buccal cell ascorbic acid concentrations were significantly lower in subjects receiving low ascorbic acid intakes (5, 10, 20mg/d) compared with repletion intakes (60 or 250mg/d) (Jacob et al., 1991). Nevertheless, the investigators concluded that buccal cells are probably not suitable as a marker of ascorbic acid status over a broad range of population intakes. In contrast, muscle tissue obtained by needle biopsy has been shown to correlate relatively well with plasma concentrations and dietary intake (Carr et al., 2013). Muscle biopsy is a highly invasive procedure and is not typically undertaken outside of research settings.

This page titled 17.12: Ascorbic acid in other cells and fluids (19.12) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.13: Body pool size (19.13)

Isotope dilution techniques can be used to determine the size of the body pool of vitamin C. The method involves administration of an oral dose of ^{14}C-labeled or ^{13}C-labeled ascorbic acid, followed by the measurement of the specific activity of blood or urine ascorbate within 24–48h (Baker et al., 1971; Kallner et al., 1977). Of the two radioactive isotopes, ^{14}C has a long half-life, so ^{13}C, a shorter-lived isotope, is preferred.

In young healthy male adults, the pool size of ascorbic acid has been estimated to be about 1500mg (i.e., 20mg/kg body weight) (Baker et al., 1971;Kallner et al., 1977). In adult men, the total body pool of vitamin C ranges in size from < 300mg to about 3000mg, depending on the daily intake of L-ascorbate. When pool sizes are < 600mg of vitamin C, psychological abnormalities have been reported (Kinsman & Hood, 1971), whereas at concentrations < 300mg, scurvy symptoms have been observed (Baker et al., 1971).

This page titled 17.13: Body pool size (19.13) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

17.14: Functional tests of vitamin C status (19.14)

There are currently no specific functional tests for vitamin C status. Bleeding into the skin to form petechia, purpura and ecchymoses are a sign of the vitamin C deficiency disease scurvy. Capillary fragility has been used in the past as a functional test of vitamin C deficiency (Vilter, 1967). However, this test can produce inconsistent results in individuals with vitamin C deficiency and is not specific to vitamin C deficiency states as other diseases may also increase capillary fragility (Vilter, 1967). Erythrocyte fragility could be an alternative functional test of deficiency (Tu et al., 2015), however, once again, this is not specific for vitamin C deficiency. Consequently, alternative functional tests of ascorbic acid status need to be developed.

Specific markers of ascorbic acid's enzyme cofactor functions have been proposed. An example relates to collagen crosslinking (Munday et al., 2005). In this study, urinary excretion of specific crosslink ratios were higher in British children versus Gambian children, and in Gambian children during the dry season (characterised by high vitamin C intake and status) versus the rainy season. A supplementation study (100mg/d of vitamin C) in Gambian children during the rainy season did not, however, alter the crosslink ratio. This may have been due to the supplementation period (7wks) being insufficient, as collagen can have a long turnover in some tissues.

Another possibility is assessment of vitamin C-dependent epigenetic marks (Young et al., 2015), for example in leukocytes. One preliminary study has demonstrated positive correlations between vitamin C status and hydroxymethylcytosine and hydroxymethyluracil marks in leukocyte DNA (Starczak et al., 2018). Participants with vitamin C concentrations > 40µmol/L exhibited significantly higher concentrations of these epigenetic marks than those with plasma concentrations < 20µmol/L. As such, more research in this area appears warranted.

This page titled 17.14: Functional tests of vitamin C status (19.14) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

18: THIAMINE (CHAPTER 20A)

Abstract

Thiamine (vitamin B_1), the first vitamin discovered and synthesized, plays a critical role in energy metabolism, neurological functioning, and early cognitive development. Thiamine deficiency disorders range broadly in presentation and severity, but beriberi can be fatal. Despite its long history, and the well-established adverse outcomes of deficiency, thiamine lacks both a standardized biomarker and interpretative criteria. The most common biomarker currently in use is blood thiamine diphosphate (ThDP), likely due to the growing availability of high performance liquid chromatography equipment and ease of standardization; however the lack of clinically meaningful cut-offs limits the practical use of ThDP. Although the functional biomarker of erythrocyte transketolase activity coefficient (ETKac) is still run in a few labs and has a well-established cut-off for deficiency, the limitations of working with washed erythrocytes and the fact that this assay is not readily available in most clinical settings hinders its use. At the individual level, use of a therapeutic thiamine challenge remains the clear course of action given the safety of thiamine, as compared to known severe adverse outcomes of untreated thiamine deficiency. At the population level, challenges still exist; however, the growing interest and method development for thiamine assessments via dried blood spots is an exciting prospect. Use of dried blood spots could ease sample collection, processing, storage, and shipment costs, widening the possibility of field-friendly assessments, for example, integrated into population-level surveys in beriberi-endemic regions. Such thiamine assessments would help enormously with properly estimating the global burden of thiamine deficiency, and triggering appropriate action for public health interventions in necessary areas.

18.1: Thiamine (20a.1)

18.2: Biochemical indices of thiamine status (20a.2)

18.3: Conclusions (20a.3)

This page titled 18: Thiamine (Chapter 20a) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

18.1: Thiamine (20a.1)

The isolation and synthesis of thiamine was carried out in 1933 by Williams and Cline (1936). The structure of thiamine consists of a pyrimidine and thiazole ring linked by a methylene bridge (Figure 20a.1). Phosphorylated forms of thiamine include thiamine monophosphate (ThMP), thiamine diphosphate (ThDP) and thiamine triphosphate (ThTP), all of which can be interconverted in the tissues. Of these, ThDP is the biologically active form and accounts for nearly 90% of total thiamine in tissues (Rindi & Laforenza, 2000).

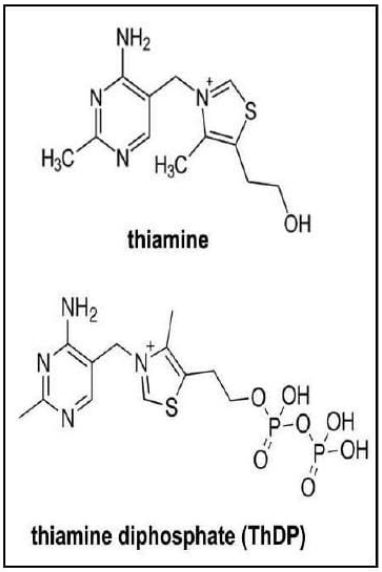


Figure 20a.1: Structure of thiamine and the coenzyme thiamine diphosphate

20a.1.1 Functions of thiamine

Thiamine, when converted to ThDP (shown in Figure 20a.1.1), has a major role in carbohydrate metabolism. As a co-enzyme in the pyruvate dehydrogenase complex, α-ketoglutarate dehydrogenase complex and the branched chain α-keto acid dehydro genase complex, ThDP is required in both the Kreb's cycle and pentose phosphate pathway (Whitfield et al., 2018). ThDP is also involved in amino acid and fatty acid metabolism. Thiamine is required for the biosynthesis of neurotransmitters (Fattal-Valevski, 2011) and in early life, thiamine has a vital role in normal cognitive development (Mimouni-Bloch et al., 2014).

20a.1.2 Deficiency of thiamine in humans

Thiamine deficiency has been classically defined as either Wernicke-Korsakoff syndrome among chronic alcoholics, or beriberi: wet beriberi for cardiac presentation, dry beriberi for neurological symptoms, Shoshin beriberi for rapid onset, or infantile beriberi for pediatric presentation. However, the broad and overlapping clinical symptoms of beriberi have recently prompted use of the terms “thiamine deficiency disorders”, or “thiamine-responsive disorders” to describe patients with symptoms or disorders that improve with thiamine treatment (Whitfield et al., 2018). Thiamine deficiency disorders have a broad clinical presentation (see examples in Table 20a.1), often leaving the disorder unrecognized, or causing mis-diagnoses (Smith et al., 2020).

Table 20a.1: Thiamine deficiency disorders (TDD) have a broad clinical presentation, with often overlapping signs and symptoms. A s of clinical presentations of TDD is summarized here. (Smith et al., 2020; Whitfield et al., 2018).

Organ system	Select manifestations of thiamine deficiency
Cardiorespiratory	* lower extremity edema (adults) or generalized edema (infants) * pulmonary hypertension * right-sided heart failure * tachycardia * tachypnea
Gastrointestinal	* anorexia * constipation * diarrhea * vomiting
Musculoskeletal	* muscle atrophy * peripheral neuropathy * weakness
Nervous	* ataxia * encephalitic presentations (brain inflammation) * impaired cognitive development (language, motor delays) * nystagmus * ophthalmoplegia

Thiamine deficiency has three general pathophysiologic mechanisms (Gomes et al., 2021):

- increased thiamine requirements, caused by fever, sepsis, malignancy, etc.;
- increased thiamine losses, caused by hemodialysis and peritoneal dialysis, chronic diuretic use, prolonged vomiting or diarrhea, etc.; and/or
- decreased thiamine intake and/or absorption, caused by malnutrition, alcoholism, hyperemesis gravidarum, consumption of thiamine antagonists or thiaminases, etc.

Thiamine deficiency often presents in low- and middle-income countries where thiamine-poor foods like white, polished rice or cassava are staples. Although historically thought to be confined to Southeast Asia, more recent reports indicate that thiamine deficiency may be a growing concern in South Asia and in other regions, such as Pacific Island nations, where white rice is increasingly consumed as a staple (Johnson et al., 2019).

It is well-established that thiamine requirements increase with higher carbohydrate intakes and greater energy expenditures (Elmadfa et al.,2001). For example, periodic outbreaks of thiamine deficiency have been reported among populations consuming monotonous rice-based diets such as illegal gold miners in French Guiana (Mosnier et al., 2017), Thai commercial fishermen (Doung-ngern et al., 2007), and Gambian farmers (Thurnham et al., 2011).

Given the role of thiamine and mechanisms of deficiency highlighted above, individuals with clinical conditions such as patients with chronic renal failure (Talwar et al., 2000), malnutrition (Hiffler et al., 2016), people living with HIV (Müri et al., 1999) and those with critical illnesses (Mates et al., 2021) may be at risk of deficiency.

The most common clinical presentation of thiamine deficiency in high-income settings is among chronic alcoholics due to general malnutrition and alcohol-induced increases in thiamine requirements and impaired thiamine absorption (Latt & Dore, 2014). Wernicke's encephalopathy is an acute, reversible presentation of thiamine deficiency that can lead to Korsakoff syndrome; the often-overlapping presentation is known as Wernicke-Korsakoff syndrome (Latt & Dore, 2014). Treatment of Wernicke-Korsakoff syndrome with both thiamine and magnesium has been recommended by Peake et al.(2013) because magnesium is also required for the full activation of the transketolase enzyme in the pentose phosphate pathway.

Another key risk group is post-operative bariatric surgery patients, in whom thiamine deficiency is caused either by lower dietary intakes or thiamine malabsorption due to the bypass or removal of the proximal small intestine (Frank, 2015). In addition, Mates et al. (2021) recently hypothesized that thiamine deficiency may be more common, yet unrecognized, among older adults in high-

income settings presenting with weakness, falling, delirium, and gastrointestinal symptoms. Thiamine deficiency in older adults could be caused by the co-existence of chronic disease burden (i.e. inflammation), insufficient thiamine intake due to age-related malnutrition, and increased thiamine losses (i.e., higher diuretic use in this population) (Mates et al., 2021). Two ongoing areas of clinical research include use of thiamine, or the highly bioavailable analogue benfotiamine, in the treatment of both Alzheimer's Disease and/or dementia (Gibson et al., 2016; Pan et al., 2016; Sambon et al., 2021) and diabetic polyneuropathy (Stracke et al., 2008).

Another growing area of interest is the impact of early life thiamine intake on cognitive development. An unfortunate “natural experiment” took place in Israel in 2003 when thiamine was erroneously excluded from infant formula, resulting in otherwise well-nourished infants presenting with symptoms of thiamine deficiency (Fattal-Valevski et al., 2005). Infants with sub-clinical infantile beriberi were followed up and later experienced delays in language(Fattal-Valevski et al., 2009) and motor skills development (Harel et al., 2017). This is relevant as the thiamine content of human milk depends on maternal thiamine intake (Dror & Allen, 2018). Thus, exclusively breastfed infants of mothers with low thiamine intakes could be at risk of either mortality via infantile beriberi, or impaired cognitive development, despite no signs of clinical deficiency. Empirical studies are currently underway to better understand the impact of thiamine exposure during the first six months of life on various aspects of cognitive and neurological development(Whitfield et al., 2019). Results thus far support that thiamine is integral to normal expressive and receptive language development (Measelle et al., 2021).

20a.1.3 Food sources and dietary intakes

Whole grain cereals, pork, and legumes are the richest food sources of thiamine, followed by other meats, fish, green vegetables, fruits and milk. Polished rice, sugar, alcohol, fat and other refined foods are poor sources of thiamine. Thiamine is thought to be highly bioavailable; notable losses would occur only with concurrent consumption of known thiamine antagonists such as betel nuts or tea, or thiaminase-containing raw fish, ferns, or African silkworm larvae (Vimokesant et al., 1982; Whitfield et al., 2018).

While whole grains are a good source of thiamine, milling induces substantial losses because most thiamine is stored in the bran; up to 50% of thiamine in wheat is lost during milling (Oghbaei & Prakash, 2016). Consequently, thiamine enrichment or fortification programs contribute greatly to thiamine intakes globally (Whitfield et al., 2021). Besides mandatory fortification of wheat flour and rice, which have been in place in Canada, the United States, and the Philippines since the 1940s (Aykroyd et al., 1945; Salcedo & Pedroche, 1949), the voluntary fortification of ready-to-eat breakfast cereals also contributes meaningfully to thiamine intakes (Fulgoni (III) & Buckley, 2015). Thiamine-rich yeast in leavened products also contributes to thiamine intake, in addition to fortified flour: Australian researchers reported that despite heat-induced baking losses, baked bread contained more thiamine than the fortified wheat flour from which it was made (Tiong et al., 2015).

Thiamine is prone to degradation by exposure to UV light, heat, oxygen and moisture (Camire et al., 1990). Of these factors, pH has the greatest effect on thiamine stability, with even weak alkaline conditions leading to thiamine degradation (Lešková et al., 2006). Thiamine is rapidly destroyed at elevated temperatures unless the pH is below 5. Hence, the addition of sodium bicarbonate to green vegetables to retain their green color destroys thiamine (Aughey and Daniel, 1940). Thiamine is lost when cooking water is discarded because it is a water-soluble vitamin. In meat thiamine is leeched into cooking juices, although losses are generally lower in dark muscles, possibly due to a protective effect from their fat content (Lešková et al., 2006). In alkaline solution, thiamine may be oxidized to the fluorescent compound thiochrome. This reaction is widely used for measuring thiamine in biological tissues and fluids.

20a.1.4 Effects of high intakes of thiamine

No adverse effects of excessive thiamine intakes have been described, except some minor transient local irritation or generalized pruritus at high treatment doses (100mg intravenous administration) (Wrenn et al., 1989). The relatively few studies conducted on the adverse effects of large doses of thiamine were insufficient to allow a Tolerable Upper Intake Level (UL) to be set by the National Academy of Medicine, formerly the Institute of Medicine (Institute of Medicine, 1998). Likewise, the European Food Safety Authority (EFSA) did not set a UL for thiamine, and concluded that the current levels of intake from thimaine from all sources do not represent a health risk for the population (EFSA 2016).

20a.1.5 Pharmacokinetics of thiamine supplementation

Oral thiamine supplementation results in a rapid increase in blood thiamine markers (Smithline et al., 2012). Thiamine is absorbed in the proximal small intestine via passive diffusion, or with an active, carrier-mediated transport mechanism at low

concentrations. Some researchers have reported that, in a single dose, no more than between 2.5 to 8.3mg is absorbed (Morrison & Campbell, 1960; Thomson & Leevy, 1972). A more recent pharmacokinetics study with 12 healthy American participants found that the thiamine absorption mechanism was not saturable with a single high oral dose of 1,500mg thiamine hydrochloride (Smithline et al., 2012). However, most hospitals would be unlikely to prescribe such a high dose; 100mg thiamine supplements are much more common (Smith et al., 2021). In a population at high-risk to thiamine deficiency in Cambodia, Coats et al. (2013)found that 16 lactating women also showed a rapid blood (and breastmilk) thiamine response to both a single 100mg dose, as well as 5 consecutive days of 100mg doses of thiamine hydrochloride.

This page titled 18.1: Thiamine (20a.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

18.2: Biochemical indices of thiamine status (20a.2)

Despite thiamine's long history, there is still debate around the best biochemical biomarker to define status (Pfeiffer et al., 2021; Whitfield et al., 2018).The two most common biochemical indices of thiamine status are the direct measure of thiamine diphosphate (ThDP) in either whole blood or erythrocytes, or the functional measure that assesses the degree of saturation of the ThDP-dependent enzyme transketolase, the erythrocyte transketolase activity coefficient (ETKac). Thiamine in erythrocytes depletes at approximately the same rate as other tissues (Brin, 1964), making blood an ideal proxy for thiamine status. Assessment of thiamine from dried blood spots is a growing research area. Other less commonly employed measures, serum/plasma thiamine and urinary thiamine excretion, will also be outlined briefly below.

Regardless of measurement (ThDP or ETKac), blood samples should be collected into blood tubes containing an anticoagulant (i.e. heparin or EDTA). Samples are stable at room temperature for a few hours, at –20°C for a few weeks, and –70°C for several months to years. Freeze-thaw cycles should be avoided (Whitfield et al., 2018). If analyses are to be completed in erythrocytes, then care must be taken to wash cells before freezing. Samples should be frozen completely (i.e. overnight at –70°C) to ensure complete lysis of erythrocytes (Lu & Frank, 2008).

20a.2.1 Thiamine diphosphate in whole blood or erythrocytes

The direct assessment of thiamine diphosphate (ThDP) in either whole blood or washed erythrocytes is currently the most common method of assessing thiamine status, even though this direct measurement does not reflect metabolic function (Whitfield et al., 2018). ThDP (sometimes along with thiamine and thiamine monophosphate (ThMP)) is most often determined via reverse-phase high performance liquid chromatography with a fluorescence detector (HPLC-FLD). A derivatizing agent (i.e. potassium ferricyanide) is used to alter the molecular structure of thiamine to thiochrome (three rings rather than two rings of thiamine); it is this three-ring structure that fluoresces and is detected. Two detailed methods are published (Lu & Frank, 2008; Mancinelli et al., 2003). Note that mass spectrometry (MS)-based methods (e.g., liquid chromatography (LC)-MS/MS) have also been developed, which allow for measurement of underivatized ThDP. However, this equipment tends to be more expensive and less commonly available than HPLC-FLD (Whitfield et al., 2018).

Erythrocyte ThDP (eThDP) and whole blood ThDP have been reported to correlate well (r=0.97) (Talwar et al., 2000), so many researchers opt to utilize whole blood samples to reduce the time and effort in sample processing of washed erythrocytes (Lu & Frank, 2008). If erythrocytes are used, cells should be washed three times in phosphate buffered saline to avoid osmotic damage to the cells (Whitfield et al., 2018), and care should be taken to remove all buffy coat as leukocytes are rich in thiamine (Chu and Hall, 1990).

Best practice is for whole blood measures to be reported in the context of hematocrit (nmol/L erythrocytes) or hemoglobin (nmol/g hemoglobin) for consistency across studies and so that whole blood and eThDP can be compared (Whitfield et al., 2018).

Interpretative criteria

There is currently no agreed upon cut-off to define risk of deficiency (Whitfield et al., 2018). Although a wide range of cut-offs have been employed, these cut-offs are often statistically derived and thus do not necessarily align with clinical signs of thiamine deficiency disorders. See Table 20a.2 for some examples of cut-offs previously used for eThDP.

Table 20a.2: Examples of eThDP cut-offs used in the literature to describe thiamine status.

Thiamine status cut-offs lower than reference range	Reference(s)	Description
< 180nmol/L	Mancinelli et al., 2003	25th percentile of n=103 (45 r
< 165nmol/L	Baines, 1985	Lower bound of 95% referenc Liverpool, UK
< 150nmol/L	Warnock, et al., 1978	Lower bound of normal range Note that this cut-off was also nmol/L

Thiamine status cut-offs lower than reference range	Reference(s)	Description
< 150nmol/L	Coats et al., 2012	Reference range of 80–150nm
< 140nmol/L	Wilkinson et al., 1997	Lower limit of normal eThDP (
< 135nmol/L	Floridi et al., 1984	eThDP reference range of 135
Marginal thiamine deficiency	Reference	Description
< 120–150nmol/L	Bailey and Finglas, 1990	Cut-off reported, but details u
Thiamine deficiency	Reference	Description
< 120nmol/L	Bailey and Finglas, 1990	Cut-off reported, but details u
< 120nmol/L	Bailey et al., 1994	eThDP < 118.5nmol/L was use

The lack of a cut-off for eThDP makes it difficult to assess deficiency among individuals. However, this is easily overcome by simply administering thiamine in a therapeutic challenge to assess thiamine responsiveness. Given the low to null risk of thiamine administration and because thiamine-deficient individuals will often respond with rapid clinical improvement within hours of treatment, thiamine responsiveness from a therapeutic dose implies deficiency (Smith et al., 2021).

At the population level, a consensus eThDP cut-off still needs to be established. This is a clear and pressing research question, as population-level risk varies widely depending on the cut-off employed. For instance, eThDP was measured in the National Micronutrient Survey linked to the Cambodian Demographic and Health Survey (2014) and suboptimal thiamine status ranged from 27% to 78% in women of reproductive age and 15% to 58% in children aged 6-69 months, depending on which cut-off was employed; see Table 20a.3 (Whitfield et al., 2017). The Global Thiamine Alliance has called for the development of a clinically relevant thiamine cut-off (Bourassa et al., 2021; Whitfield et al., 2018).Comparison of ETKac and eThDP revealed an inflection point at an ETKac of 1.25 (indicating deficiency) with a ThDP of 270 ng/g hemoglobin (Talwar et al., 2000). This converts to approximately 224nmol ThDP/g hemoglobin. However, this comparison was made among a relatively small group of 63 medical and surgical patients in the United Kingdom thought to be at risk of thiamine deficiency (Talwar et al., 2000). Hence, it is unclear whether this cut-off would be valid among populations with endemic thiamine deficiency (Whitfield et al., 2018).

Table 20a.3: Erythrocyte ThDP and the application of various thiamine status eThDP cut-offs from the literature, among women and *2017).*

[a] *See Table 20a.2 for more information about how the thiamine status cut-offs in this table were derived.*

Thiamine status eThDP, nmol/L	Women aged 15–49y, n=719
Mean (95% CI)	150 (146–153)
Range	41–352
Thiamine status cut-offs[a]	
Lower than reference range	
< 180nmol/L	558 (78%)
< 165nmol/L	481 (67%)
< 150nmol/L	398 (55%)
< 140nmol/L	336 (47%)
< 135nmol/L	292 (41%)
Marginal thiamine deficiency	
120 – 150nmol/L	208 (29%)
Thiamine deficiency	
< 120nmol/L	192 (27%)

Factors affecting whole blood or erythrocyte ThDP

Concern has been raised about the use of the direct assessment of ThDP in blood samples in populations with persistently low thiamine intakes, such as groups in low- and middle-income countries heavily reliant on white, polished rice and likely also at risk of chronic inflammation and/or frequent infection. This concern is highlighted in a case-control study in rural Cambodia that found no differences in whole blood ThDP (corrected for hematocrit) between breastfed infants with and without clinical signs of beriberi and the mothers of the cases and controls (Coats et al., 2012). The means (95% CI) ThDP/Hct (i.e.,corrected ThDP) for infant cases and controls were 135 (114–157) and 159 (133–184nmol/L) erythrocytes (p=0.08), respectively. For mothers of cases and controls, authors reported 141 (126–158) and 150 (124–166nmol/L) erythrocytes (p=0.46), respectively (Coats et al., 2012). These results call into question whether ThDP is a valid biomarker of thiamine status.

20a.2.2 Erythrocyte transketolase activity coefficient (ETKac)

Transketolase (EC 2.2.1.1) is a ThDP-dependent enzyme. Measurement of the activity of this enzyme in erythrocytes is commonly used as an index of thiamine nutritional status. The principle of this assay is to assess the saturation of the ThDP-dependent enzyme transketolase by measuring enzyme activity before and after the addition of excess exogenous ThDP.

In outline, this assay and similar procedures are used for parallel assays for riboflavin and pyridoxine and involves the following stages:

- The basal activity of the enzyme (transketolase) in erythrocytes is measured. This represents the endogenous enzyme activity and depends on the amount of the coenzyme (ThDP) in the erythrocytes.
- The enzyme activity with excess coenzyme added *in vitro* is then determined. This equates to the maximum potential enzyme activity and is referred to as total or “stimulated” activity.
- The basal and enzyme activities are then compared to indicate the degree of unsaturation of the enzyme with the coenzyme.

Transketolase catalyzes the following two reactions in the pentose phosphate pathway:

$$\begin{array}{c} \text{xylulose-5-phosphate} + \text{ribose-5-phosphate} \\ \leftrightharpoons \\ \text{sedoheptulose-7-phosphate} + \text{glyceraldehyde-3-phosphate} \end{array}$$

and

$$\begin{array}{c} \text{xylulose-5-phosphate} + \text{erythrose-4-phosphate} \\ \leftrightharpoons \\ \text{fructose-6-phosphate} + \text{glyceraldehyde-3-phosphate} \end{array}$$

The results of this assay are typically expressed in terms of the activity coefficient (ETKac), calculated as a ratio of the enzyme activity (with excess coenzyme) divided by the basal enzyme activity (without coenzyme). When the vitamin status is adequate, the added coenzyme has little effect on the total enzyme activity, so the activity coefficient is very close to 1.0. However, in vitamin deficiency states, the added coenzyme increases the enzyme activity. Larger increases are associated with higher values for the activity coefficient and thus greater degrees of deficiency. This is shown diagramatically in (Figure 20a.2).

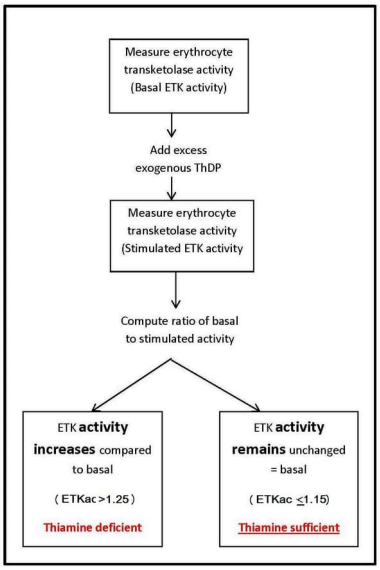


Figure 20a.2: Diagram of the principle of the erythrocyte transketolase activity coefficient assay

Sometimes the percentage of stimulation, also known as the percentage activation α, is calculated from the activity coefficient as (ETKac × 100) – 100. This is known classically as the thiamine pyrophosphate (diphosphate) effect, “TPP effect”. If not reported as a ratio (ETKac), the basal and stimulated enzyme activities can be expressed per gram of hemoglobin, per number of erythrocytes, or in terms of the volume of erythrocytes (in mL).

Based originally on methods of Brin (1966) and later Smeets et al. (1971)and Vuilleuimier et al. (1983), this method has been recently updated for analysis on a plate reader with a more convenient 96-well plate configuration (Jones et al., 2021). An important consideration for this 96-well format in particular is the need for identical temperature exposure to all wells: the plate reader's incubator must be functioning well to maintain analytical temperatures (Whitfield et al., 2018).

ETKac Interpretive criteria

The interpretative criteria for ETKac, shown in Table 20a.4, are consistent throughout the literature (Institute of Medicine, 1998; Ruston et al., 2004; Turck et al., 2016; World Health Organization, 1999).

Table 20a.4: Interpretative criteria for erythrocyte transketolase activity

ETKac and α activation	Status
ETKac < 1.15, or α < 15%	adequate thiamine status
ETKac 1.15–1.25, or α 15–25%	thiamine insufficiency / marginal thiamine deficiency
ETKac > 1.25, or α > 25%	thiamine deficiency

Factors affecting erythrocyte transketolase activity

In thiamine deficiency, the basal level of erythrocyte transketolase activity is low and an enhancement in the enzyme activity after the addition of ThDP *in vitro* is generally observed. Prolonged thiamine deficiency, however, induces a reduction in the apotransketolase enzyme level. Consequently, both the basal and stimulated erythrocyte transketolase activities tend to fall, with no change in the percentage activation α, giving rise to misleadingly “normal” ETKac values (Schrijver, 1991). In such cases, the basal enzyme activity should also be considered (Nixon et al., 1990; Soukaloun et al., 2011).

Use of the ETK ratio versus basal activity. The ratio of stimulated to basal enzyme activity is used as the primary measure because (a) the inter-subject variation in basal erythrocyte enzyme activity is large, and (b) it is assumed that apoenzyme levels are not affected by vitamin deficiencies. This latter assumption may not be correct, however (Schrijver, 1991). Vitamin deficiency

or excess and other factors such as the presence of certain diseases and the administration of hormones and drugs, may affect apoenzyme levels and confound the interpretation (Talwar et al., 2000). Hence, when interpreting the results, it is advisable to take into account the basal enzyme activity, as well as its activation with coenzyme (Bamji, 1981). Jones et al. (2020) argue that use of the ratio reduces the need for the precise assay conditions such as the optimal path length, which is required when expressing basal and stimulated activities alone (Bates, 1997). Conversely, Soukaloun et al. (2011) specifically recommend interpretation of basal ETK rather than the ratio ETKac among infants. They argue that, due to low thiamine exposure *in utero* and postnatally, apoenzyme may be unstable and concentrations may fall. As a result, ETKac could underestimate thiamine deficiency in infancy.

Age of the erythrocytes influences enzyme activity; levels decline as the cells age. As a result, the basal enzyme activity depends on the mean age of the erythrocytes (Thurnham, 1981). Hence, in patients undergoing treatment for iron deficiency, for example, the basal level of erythrocyte transketolase activity will be increased as a result of the reticulocytosis (Thurnham, 1981).

Age affects transketolase activity. Older individuals tend to have lower activities, possibly as a result of a defect in the apoenzyme rather than inadequate activity of the coenzyme (O'Rourke et al., 1990).

Certain disease states are associated with low erythrocyte transketolase activity (Talwar et al., 2000). In some cases (e.g., diabetes and liver disease), the reduced activity results from a reduced apoenzyme level associated with the disease. In other diseases (e.g., certain cancers), the percentage activation α is enhanced despite adequate thiamine intakes because the conversion of thiamine to ThDP is impaired (Basu & Dickerson, 1976). Patients with polyneuritis, uremic neuropathy, and disorders of the gastrointestinal tract, also have low transketolase activity (Kjøsen & Seim, 1977), although the cause is uncertain.

Intra-subject variation in erythrocyte transketolase activity can be large, even in individuals with fixed thiamine intakes (van Dokkum et al., 1990).Such large variations occur because individuals differ in their sensitivity to thiamine deficiency. This leads to difficulties in interpreting transketolase activity coefficients when values are indicative of adequacy (i.e., 1.0) and those indicative of thiamine deficiency (i.e., ≥ 1.25). Hence, correlation of ETKac and intakes within this range may be poor (Vir & Love, 1979).

In view of these uncertainties, it is perhaps not surprising that ETKac may not correlate significantly with thiamine intakes. In general, such relationships have only been reported in surveys of population groups known to be vulnerable to thiamine deficiency, such as the elderly. Correlations, however, are less obvious in groups such as adolescents who are more likely to have adequate thiamine intakes (Bailey et al., 1994). More recent studies that have measured transketolase activity generally do not report correlations between dietary thiamine and transketolase activity (2011). An exception is a study in Laos, where postpartum restrictive diets are common, when Khounnorath et al. (2011) surpisinly reported lower basal transketolase activity among infants of mothers who had eaten pork since delivery, a good source of thiamine. Although a number of studies have found increased whole blood or erythrocyte ThDP concentrations after thiamine supplementation (e.g., (Coats et al., 2013; Smithline et al., 2012), few studies have examined ETKac in the context of thiamine supplementation. Gallant et al. (2021) recently reported significantly lower ETKac among lactating women consuming a daily, low-dose thiamine supplement (between 1.2 and 10mg/d) over 22 weeks, as compared to a placebo; also see Table 20a.5 However, higher dose supplementation (100mg/d) among 25 Karen mothers in the Mae La refugee camp on the Thai-Myanmar border did not significantly impact ETKac when compared to the response of 22 unsupplemented mothers (McGready et al., 2001).

Likewise, abnormal erythrocyte transketolase activity may not be associated with clinical signs of thiamine deficiency disorders. For example, in the same study of Karen refugees, symptoms such as paresthesia were not associated with abnormal erythrocyte transketolase activity at 30 wk gestation or at delivery (McGready et al., 2001). However, both basal transketolase activity and ETKac differed significantly between infants with clinical beriberi (n=47) and controls (n=47 febrile controls and n=47 afebrile controls) in a Laos study (2011).

The ETKac assay is often critiqued for poor inter-assay precision, as well as a lack of standardization (Whitfield et al., 2018) and challenges with inter-laboratory comparisons (Bailey et al., 1994). More recently, the limited availability of the ETKac assay has prevented widespread use, particularly in clinical, remote, or low-income settings (Barennes et al., 2015; Sastry et al., 2021). It appears that only a few laboratories in Scotland and England still actively assess ETKac (Jones et al., 2021; Maguire et al., 2021). Two examples of ETKac from a low- and high-risk population, without and following thiamine supplementation are shown in Table 20a.5.

Table 20a.5: Recent examples of ETKac from a low-risk United Kingdom population sampled during the National Diet and Nutrition S population of rural, lactating women 18–45y participating in a 22-week thiamine supplementation research study (Gallant et al., 202

Population	Participant group	n
United Kingdom	4–10y	101

Population	Participant group	n
	11–18y	164
	19–64y	513
	≥ 65y	153
Cambodia	2 weeks postpartum, before any thiamine supplementation	334
	24 weeks postpartum (placebo group)	83
	24 weeks postpartum, after 22 weeks of thiamine supplementation: 1.2mg/d	86
	24 weeks postpartum, after 22 weeks of thiamine supplementation: 2.4mg/d	81
	24 weeks postpartum, after 22 weeks of thiamine supplementation: 10mg/d	85

20a.2.3 Thiamine assessment with dried blood spots

Although blood is an ideal medium for assessing thiamine status (Brin, 1964; Whitfield et al., 2018), there are several challenges with venipuncture, particularly in low-income settings. It is an invasive procedure that requires clinical expertise (trained phlebotomists), facilities to process samples (refrigerated centrifuge if washing erythrocytes) and a cold chain Huang et al., 2020; Verstraete & Stove, 2021b). For infants in particular, the volume of blood, invasiveness of venipuncture, and the need to hire pediatric phlebotomists, may lead to challenges with traditional venous blood collection. The utilization of dried blood spots addresses many of these challenges, simplifying both sample collection and storage. Dried blood spots would allow fingerprick capillary blood collection to replace venipuncture and there is potential for ambient, rather than −80°C freezer, storage (Verstraete & Stove, 2021b). Even if cold chain transport is required, dried blood spots are much smaller and pose a lower biohazard risk, than aliquots of frozen blood.

A few different methods employing dried blood spot sample collection have been published in recent years. Mathew et al. (2019) validated a method using HPLC-FLD. However, the method requires a known volume of venous blood pipetted onto the filter paper cards, limiting its practical use. Huang et al. (2020) also developed a HPLC-FLD method, while Verstraete and Stove's method uses LC-MS/MS (Verstraete & Stove, 2021a).

In 2020, a method to assess ThDP, as well as thiamine and ThMP from capillary dried blood spots was developed by Huang et al.(2020). Authors collected venous whole blood samples (stored at −80°C until analysis), as well as capillary fingerprick samples (stored at −20°C until analysis), from 20 healthy Australian volunteers. There was a strong linear correlation between the results for venous whole blood ThDP and capillary dried blood spot ThDP ($r=0.9636$; $p< 0.001$). ThDP coefficients of variation (%CVs) for low and high quality controls were: intra-day: 3.5% and 2.7%, respectively; inter-day: 4.3% and 3.0%, respectively. Nevertheless, the ThDP measured in capillary dried blood spots was 6.4% higher as compared to venous whole blood, which authors speculated could be attributed to the higher hematocrit values of capillary samples. Hence, although promising, this method has a few limitations. While sample extraction from the filter paper cards was completed with a standard 6.35mm punch disc, a practice intended to minimize the need for hematocrit assessment, a limitation of this method is the unknown sample volume. In addition, cold chain storage is likely to still be required for this method: although ThDP appeared to be stable when dried blood spot cards were stored at −20°C, preliminary analyses showed significant losses during ambient storage (Huang et al., 2020).

As noted above, a standard sample volume or hematocrit assessment is essential because ThDP is found mainly in erythrocytes (Whitfield et al., 2018). However, Verstraete and Stove recently reported a novel method employing volumetric absorptive microsampling (VAMS), wherein a collection device with a plastic handler and polymeric tip absorbs a fixed volume of blood, thus negating the need for additional hematocrit assessment(Verstraete & Stove, 2021b). Authors reported good agreement between venous whole blood and capillary VAMS samples, with a mean bias of −1.0% (95% CI: −4.1%, 2.0%) which was not significantly different from zero. The %CVs were 6.5% and 2.2% for capillary VAMS and venous whole blood samples, respectively. “Uncontrolled transport” was also assessed by sending capillary VAMS samples through the regular mail, thus subjecting samples to 2–5d of ambient conditions. These “uncontrolled” samples had a mean difference of 1.0% (95% CI: −18.6%, 20.6%) and 1.0% (95% CI: −14.7%, 16.7%) in two analyses when compared to “controlled” capillary VAMS samples that had been stored at −80°C until analysis. Although only 3% of samples exceeded a 20% difference threshold set to compare “controlled” and “uncontrolled” samples, the wide 95% CI is of concern and cold chain transport and storage may still be warranted. This is particularly true as these “controlled” samples were mailed during the Belgian winter, when ambient temperatures were likely low (Verstraete & Stove, 2021b). However, during method development Verstraete and Stove reported that VAMS samples were stable for 1 month at room temperature, or for 1 week at either 60°C or at 80% humidity (Verstraete & Stove, 2021a).

A strength of this method is the vastly simplified sample collection and known sample volume. Although promising, this method comparison was completed among 50 healthy Belgian volunteers with relatively high ThDP concentrations. As such, method assessment among participants with lower thiamine status and with health profiles more common in beriberi- endemic regions (e.g. higher inflammation, genetic hemoglobin disorders, environmental enteric dysfunction, malaria, tuberculosis, HIV, etc) should be undertaken.

Interpretive criteria

To date, published work has focussed on method development and comparisons; no thiamine assessments have been completed at large scale with dried blood spot samples. As a result, appropriate cut-offs have not yet been established. Although the same HPLC-FLD or LC-MS/MS methods are employed as those used for venous samples, the established differences between venous and capillary blood should be taken into account when a cut-off is established. In both infants and adults, capillary blood samples have higher hematocrit and red blood cell counts than venous samples(Kayiran et al., 2003), so that cut-offs may need to be lower if capillary samples are collected.

20a.2.4 Serum or plasma thiamine

Serum or plasma thiamine assessments with either HPLC-FLD or LC-MS/MS are available in both clinical (Mates et al., 2021) and research settings(McCann et al., 2017). McCann et al. (2017) compared plasma total thiamine (thiamine + ThMP) to eThDP among 196 rural Cambodian women of reproductive age before and after a 6mos *ad libitum* thiamine-fortified fish sauce intervention. They found a moderate, significant correlation between measures (baseline: ρ=0.554, $p < 0.01$; post-thiamine intervention: ρ=0.60, $p < 0.01$) (McCann et al., 2017). However, the use of serum or plasma thiamine concentrations to assess thiamine status is generally not advised, as these assessments often measure free thiamine or ThMP, which are not reflective of total body stores (Mates et al., 2021). In addition, serum or plasma would only reflect recent intake rather than true tisssue thiamine status (Whitfield et al., 2018). Serum or plasma thiamine could also be impacted by various disease states such as acute hepatic injury and is thought to be transiently decreased in critically ill patients after trauma or those experiencing sepsis (Frank, 2015).

20a.2.5 Urinary thiamine excretion

Urinary thiamine is no longer a common method of thiamine assessment. Collection of 24h urine samples is logistically challenging and imposes a high participant burden. Although casual urine samples can be collected, when thiamine excretion is expressed per g creatinine, urinary thiamine is still not the best indicator of status. For instance, urinary thiamine excretion did not correlate well with ETKac, or with clinical signs of deficiency among a group of 79 alcoholic patients in Australia (Wood et al., 1977).

In principle, among thiamine-replete individuals, when the main tissue compartment is saturated and the urinary excretion threshold is exceeded, thiamine levels in the urine do not generally reflect body stores; instead, they parallel recent dietary intakes. In contrast, when the tissues are unsaturated, excretion is limited because the body conserves thiamine. When dietary intakes are sufficient, for example 0.5mg thiamine per 1,000kcal, approximately 100µg thiamine is excreted (Sauberlich, 1967). At lower intakes, however, urinary excretion of thiamine is no longer linearly related to intake (Oldham, 1962). This critical point is thought to be around intakes of 0.2mg thiamine per 1,000kcal, yielding urinary thiamine excretions between 5–20µg thiamine, with clinical beriberi cases yielding excretions of 0–15µg thiamine (Sauberlich, 1967). In addition, at low thiamine intakes, most of the thiamine excretion is in the form of metabolites such as pyramin, an even less sensitive measure than urinary thiamine (Sauberlich, 1967).

The intra-subject variability for urinary thiamine excretion based on a single 24h urine sample is high, with %CVs ranging from 11% to 13%, even for participants on a constant diet for 60d (van Dokkum et al., 1990). Furthermore, urinary thiamine excretion is increased by the use of certain drugs, particularly diuretics (Suter & Vetter, 2000). In addition, age-specific interpretive criteria are likely to be necessary as children have higher levels of thiamine excretion than adults when expressed on a per g creatinine basis (van Dokkum et al., 1990).

This page titled 18.2: Biochemical indices of thiamine status (20a.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

18.3: Conclusions (20a.3)

Despite its long history and the well-established adverse effects of deficiency, thiamine lacks both a standardized biochemical biomarker and interpretative criteria. The most common biomarker currently in use is blood ThDP, likely due to the growing availability of HPLC-FLD and ease of standardization. However, the lack of clinically meaningful cut-offs limits the practical use of blood ThDP. Although ETKac is still measured in a few laboratories and has a well-established cut-off for deficiency, the lack of availability of the assay limits its use. At the individual level, use of a therapeutic thiamine challenge remains the clear course of action given the safety of thiamine, when compared to the known severe adverse outcomes of untreated thiamine deficiency. At the population level, challenges still exist. However, the growing interest and method development for thiamine assessment via dried blood spots is an exciting prospect. Use of dried blood spots could ease sample collection, processing, storage and shipment costs, thus extending the possibility of field- friendly assessments, such as integration into population-level surveys in beriberi-endemic regions. Such thiamine assessments would provide improved estimates of the global burden of thiamine deficiency and trigger appropriate action for public health interventions in necessary areas. Future research should continue exploring: 1) large-scale testing of dried blood spot samples and potentially also the development of a point-of-care thiamine assessment; 2) the development of interpretative criteria for whole blood ThDP; and 3) improved inter-laboratory and standardization of ETKac.

This page titled 18.3: Conclusions (20a.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

19: RIBOFLAVIN (CHAPTER 20B)

Abstract

Riboflavin is a component of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) which act as coenzymes of different flavoproteins involved in oxidation-reduction reactions that are central to energy production, intermediary metabolism, drug metabolism and the maintenance of antioxidant status. Riboflavin coenzymes are also required for the metabolism of folate, vitamin B_{12}, vitamin B_6, and niacin as well as for mobilization of iron from intracellular stores. Classical signs of riboflavin deficiency (ariboflavinosis) generally occur in association with other nutrient deficiencies, notably vitamin B_6. Riboflavin deficiency has been described predominantly in undernourished populations in low income countries, but accumulating evidence has shown that a suboptimal riboflavin status is more common than previously recognized among populations in developed countries. Measurement of the activity of glutathione reductase, with and without the prosthetic group FAD, is the best method for assessing tissue riboflavin status, particularly in cases of impaired riboflavin status. The test measures tissue saturation and long-term riboflavin status, although some confounding factors may influence its performance. Urinary riboflavin excretion levels in casual or 24h urine specimens reflect dietary intake but vary widely because concentrations are affected by many non-nutritional factors. Concentrations of riboflavin, FMN and FAD in plasma or erythrocytesG as well as measurement of the activity of pyridoxamine phosphate oxidase and its activity coefficient appear promising options for assessment of riboflavin status and warrant further exploration.

19.1: Riboflavin (20b.1)

19.2: Biomarkers of riboflavin status (20b.2)

This page titled 19: Riboflavin (Chapter 20b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

19.1: Riboflavin (20b.1)

Riboflavin (7,8-dimethyl-10-ribityl-isoalloxazine) was first synthesized in 1935. Its structure consists of an isoalloxazine ring attached to a ribityl side chain Figure 20b.1. A detailed review of riboflavin and health is provided elsewhere (Powers, 2003; Thakur et al., 2017; Suwannasom et al., 2020).

20b.1.1 Functions of riboflavin

Riboflavin is a component of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) which act as coenzymes of different flavoproteins involved in oxidation-reduction reactions that are central to energy production, intermediary metabolism, drug metabolism as well as the maintenance of antioxidant status (Figure 20b.1). Of the two coenzymes, FMN is formed first, from free riboflavin, by ATP-dependent phosphorylation; a reaction catalyzed by cytosolic flavokinase (EC 2.7.1.26). Next, most of this FMN is combined with a molecule of ATP to form FAD; this step is catalyzed by the FAD-dependent FAD synthetase (EC 2.7.7.2). In turn, FAD can be converted into forms covalently bound to tissue proteins. Thyroid hormones regulate the synthesis of FMN and FAD as well as the formation of the covalently bound flavins (Rivlin, 1970; Pinto and Rivlin, 1979).

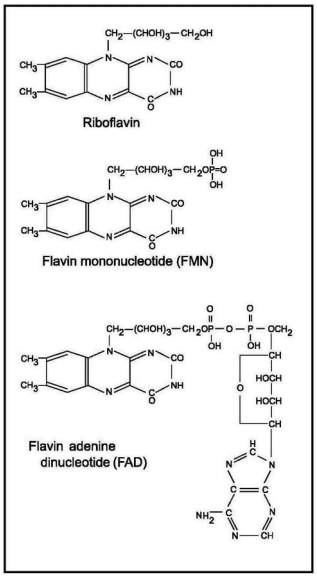


Figure 20b.1: Structure of riboflavin and the two coenzymes derived from flavin, flavin mononucleotide, and flavin adenine dinucleotide.

Both FMN and FAD are also cofactors for several enzymes, including glutathione reductase, xanthine oxidase, L-amino oxidase, and nicotinamide adenine dinucleotide (NAD) dehydrogenase. Results from animal studies also suggest that flavins are cofactors in the cyclical β-oxidation of fatty acids (Olpin and Bates, 1982a; Olpin and Bates, 1982b). Around 84% of human flavoproteins are FAD-dependent and only 16% use FMN as a cofactor (Leinhart et al., 2013).

Riboflavin coenzymes are also involved in the metabolism of four other vitamins — folate, vitamin B_{12}, vitamin B_6, and niacin. Flavin adenine dinucleotide is a cofactor for 5,10-methylenetetrahydrofolate reductase (MTHFR; EC 1.7.99.5), a key enzyme in folate metabolism involved in the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, which is required for the remethylation of homocysteine to methionine. Both FMN and FAD act as cofactors for the enzyme methionine synthase reductase (EC 1.16.1.8) which is responsible for the regeneration of methylcobalamin, the biologically active form of vitamin B_{12} that is also involved in the remethylation of homocysteine to methionine. In addition, FMN acts as a cofactor for pyridoxine

(pyridoxamine) phosphate oxidase (PPO; EC 1.4.3.5), an important enzyme in vitamin B_6 metabolism that converts the 5-phosphates of both pyridoxine and pyridoxamine to the coenzyme pyridoxal-5´-phosphate. Finally, the FAD-dependent kynurenine hydroxylase is involved in the conversion of tryptophan to niacin. As a result, severe riboflavin deficiency can cause functional deficiency with disturbances in the metabolic pathways of these vitamins.

Absorption of riboflavin occurs mainly in the proximal small intestine by a saturable carrier-mediated mechanism; the maximal amount of riboflavin that can be absorbed from a single dose is about 27mg (Zempleni et al., 1996). Intestinal microflora is able to synthesize riboflavin, some of which is absorbed in the colon by a carrier-mediated process (Said et al., 2000; Said, 2013). The vitamin is transported in plasma as free riboflavin complexed with albumin and other proteins, mainly immunoglobulins, which also bind flavin coenzymes (Innis et al., 1985).

20b.1.2 Deficiency of riboflavin in humans

The classical signs of riboflavin deficiency, termed ariboflavinosis, are angular stomatitis, cheilosis, glossitis and anemia. Corneal vascularization, dermatological changes, and neurological alterations may also occur but are not specific for ariboflavinosis (Northrop-Clewes and Thurnham, 2012). Some environmental factors may also influence the clinical signs and symptoms of riboflavin deficiency.

Ariboflavinosis is rarely encountered in isolation, usually occurring in association with other vitamin deficiency states. In a study of Irish elderly, 49% had suboptimal riboflavin status, 39% had suboptimal vitamin B_6status, and 21% had concurrent riboflavin and vitamin B_6 deficiencies. After supplementation with riboflavin alone, both the riboflavin deficiency and the low plasma pyridoxal-5´-phosphate levels were corrected (Madigan et al., 1998).

Some evidence suggests that riboflavin deficiency is linked with impaired mobilization of iron from intracellular stores (i.e. hepatic ferritin), increased rate of iron loss from the gastrointestinal tract, as well as decreased absorption of iron (Powers, 2003). This could explain the low hemoglobin concentrations and hypochromic anemia which often are observed in riboflavin deficiency (Aljaadi et al., 2019). Studies conducted in population groups with a compromised riboflavin status such as school children (Buzina et al., 1979; Charoenlarp et al., 1980), men (Fairweather-Tait et al., 1992), pregnant (Decker et al., 1977; Suprapto et al., 2002; Ma et al., 2008), lactating (Powers et al., 1985) and women of reproductive age (Powers et al., 2011), have shown that riboflavin supplementation leads to improvement of the hematological status.

Riboflavin deficiency has been described in undernourished populations in several low income countries, notably among women and children in The Gambia (Bates et al., 1981, 1994) , some elderly persons in Guatemala (Boisvert et al., 1993a), children in Côte d'Ivoire (Rohner et al., 2007), and in some adolescent refugees from Bhutan living in southeastern Nepal (Blanck et al., 2002). Infants of mothers with a low riboflavin status during gestation are also likely to be born riboflavin-deficient (Bates et al., 1982). An accumulating body of evidence based on small as well as population-based surveys from developed countries has shown that a suboptimal riboflavin status is more common than previously recognized and those predominantly affected are adolescents and young women (Bates et al., 2016; Ward et al., 2020; Aljaadi et al., 2019; Jungert et al., 2020).

Several conditions, including alcoholism, diabetes mellitus, liver disease, thyroid and adrenal insufficiency, and gastrointestinal and biliary obstruction, may precipitate or exacerbate riboflavin deficiency (Rivlin, 2007). Alcohol causes deficiency by interfering with both the digestion and the intestinal absorption of riboflavin S (Pinto et al., 1987; Subramanian et al., 2013). Antiepileptic and psychotropic drugs such as chlorpromazine, imipramine, and amitriptyline (Apeland et al., 2003; Pinto et al., 1982), as well as some antimalarial drugs such as quinacrine (Dutta et al., 1985) all inhibit the conversion of riboflavin to its active coenzyme derivatives. Drugs, such as tetracycline, theophylline, and caffeine, as well as metals, such as zinc, copper, and iron, may chelate or form complexes with riboflavin and, hence, affect its bioavailability (Sauberlich, 1985).

More vulnerable to riboflavin deficiency are individuals with 677C→T polymorphism in the gene encoding MTHFR enzyme because the variant enzyme has an impaired activity as a result of reduced affinity to the cofactor FAD (Yamada et al., 2001); this typically leads to elevated homocysteine concentrations (Frosst et al., 1995). Marked lowering of homocysteine has been achieved in people with the impaired variant (TT genotype) with a low dose of riboflavin (McNulty et al., 2006). Moreover, genome-wide association studies (Ehret et al., 2011) and clinical studies (Qian et al., 2007; Yang et al., 2014) provide evidence linking the MTHFR 677C→T polymorphism with blood pressure and increased risk of hypertension and hypertension in pregnancy by up to 87%. Importantly, emerging evidence from randomized trials highlights that riboflavin supplementation can lower blood pressure specifically in adults with TT genotype (Horigan et al., 2010; Wilson et al., 2013). This novel role of riboflavin could have important public health implications, considering that the frequency of the variant TT genotype is 10–12% worldwide but it could reach up

to 32% in some countries (e.g., Mexico; Wilcken et al., 2003).

Riboflavin deficiency has been implicated as a risk factor for cancer in animal studies (Powers, 2003), but the epidemiological evidence based on several meta-analyses is less consistent (Yoon et al., 2016; Yu et al., 2017; Ben et al., 2019; Zeng et al., 2020).

Studies have shown that in malaria endemic regions, individuals with riboflavin deficiency are relatively resistant to this infection and have a lower level of parasitemia, but the course of the disease may be more severe than in people with adequate riboflavin status (Das et al., 1988).

20b.1.3 Food sources and dietary intakes

Riboflavin in foods is present predominantly in the form of FAD and only small quantities are available as FMN. Some flavins bound covalently to protein are also found in certain foods but are largely unavailable as nutritional sources of riboflavin; only limited amounts apparently undergo digestion and absorption (Chia et al., 1978).

The major food sources of riboflavin are dairy products, especially milk, and meat and fish; most plants contain only small amounts of riboflavin. As a result, individuals consuming mainly plant-based diets may be at risk for riboflavin deficiency. Riboflavin enrichment of flour is mandated in the US and some other countries with the aim of restoring the losses of the vitamin during milling and refining processes. Most of the ready-to-eat breakfast cereals are fortified with riboflavin in developed countries. Although it is heat-stable, losses of riboflavin do occur if it is exposed to the light and, as it is water soluble, by leaching into the cooking water (Powers, 2003). Bioavailability of riboflavin from food is reported to be around 95% (IOM, 2000). There are limited data on the relative bioavailability of riboflavin from different food sources but a study using stable isotopes and kinetic modeling did not find a significant difference in riboflavin absorption from milk and spinach (Dainty et al., 2007).

Nationally representative surveys from the US, Australia, Ireland, and the UK showed that milk and dairy products, meat and ready-to-eat breakfast cereals are the main dietary contributors to riboflavin intake (IOM, 2000; Australian Bureau of Statistics, 2014; National Adult Nutrition Survey, 2011; Bates et al., 2016).

20b.1.4 Effects of high intakes of riboflavin

In humans, there is no evidence for riboflavin toxicity as a result of excessive intakes (EFSA, 2017). The absorption of orally administered riboflavin from both vitamin supplements and from natural foodstuffs is limited, and high intakes are rapidly excreted in the urine. Even when 400mg/d of riboflavin was given orally with meals for at least 3 months, no short-term side effects were reported (Schoenen et al., 1994). In view of the limited data on adverse effects from high intakes of riboflavin, no Tolerable Upper Intake Level for riboflavin was set by the U.S. Food and Nutrition Board (IOM, 2000).

This page titled 19.1: Riboflavin (20b.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

19.2: Biomarkers of riboflavin status (20b.2)

As the signs and symptoms of riboflavin deficiency are not very specific, diagnosis of a deficiency state is difficult when based exclusively on clinical assessment. Consequently, biochemical tests are essential for confirming clinical cases of riboflavin deficiency and for establishing subclinical deficiencies. Several tests are available, and these are discussed in the following sections. A combination of biomarkers is generally preferred: erythrocyte glutathione reductase activity and urinary riboflavin excretion are the most used.

20b.2.1 Erythrocyte glutathione reductase activity

The measurement of the activity coefficient of erythrocyte glutathione reductase (EGR) (EC 1.6.4.1), an erythrocyte enzyme that depends on a cofactor derived from riboflavin, is the preferred method for assessing riboflavin status. It provides a measure of tissue saturation and long-term riboflavin status.

Glutathione reductase is a nicotinamide adenine dinucleotide phosphate (NADPH), a FAD-dependent enzyme, and the major flavoprotein in erythrocytes. It catalyzes the oxidative cleavage of the disulfide bond of oxidized glutathione (GSSG) to form reduced glutathione (GSH):

$$\mathrm{GSSG} + \mathrm{NADPH} + \mathrm{H}^{+} \rightarrow 2\mathrm{GSH} + \mathrm{NADP}^{+}$$

The activity of EGR is measured spectrophotometrically by monitoring the oxidation of NADPH to $NADP^{+}$ at 340nm, with and without the presence of added FAD coenzyme. As an alternative, the production of GSH can be monitored colorimetrically. The test involves the following steps:

- The basal activity of EGR is measured. This represents the endogenous enzyme activity and depends on the amount of FAD coenzyme in erythrocytes.
- The activity of EGR with excess FAD coenzyme added *in vitro* is then determined. This represents the maximum potential EGR activity and is referred to as “stimulated” activity.
- The activity coefficient (EGR AC) is then derived by the ratio of the “stimulated” over the basal EGR activity which indicates the degree of unsaturation of the enzyme with the coenzyme:

$$\mathrm{EGRAC} = \frac{\text{activity (with added FAD)}}{\text{basal activity (without added FAD)}}$$

The basal and stimulated EGR activities can be expressed per gram of hemoglobin, per number of erythrocytes, or in terms of the volume of erythrocytes (in mL).

The degree of *in vitro* stimulation of EGR activity depends on the FAD saturation of the apoenzyme, which, in turn, depends on the availability of riboflavin. If the vitamin status is normal, then the values of the basal and stimulated activity are similar, and the EGR activity coefficient is very close to 1.0. In persons with riboflavin deficiency, the basal EGR activity falls and the *in vitro* stimulation by FAD rises; thus, higher values of EGR activity coefficient are indicative of a lower riboflavin status. The result also can be presented as “percentage of stimulation” (PS) which is calculated from the EGR AC as follows:

$$\mathrm{PS} = (\mathrm{EGRAC} \times 100) - 100$$

Meta-analysis of human supplementation trials confirmed that the EGR AC is a useful, stable, and sensitive measure of riboflavin status, reflecting the degree of tissue saturation ranging from severe deficiency to normal status (Hoey et al., 2009). It is not very sensitive, however, for detecting changes of riboflavin status in replete individuals (Powers, 1999). In experimentally controlled human studies, concomitant increases in EGR AC in response to decreases in intakes of riboflavin have been reported (Figure 20b.2).

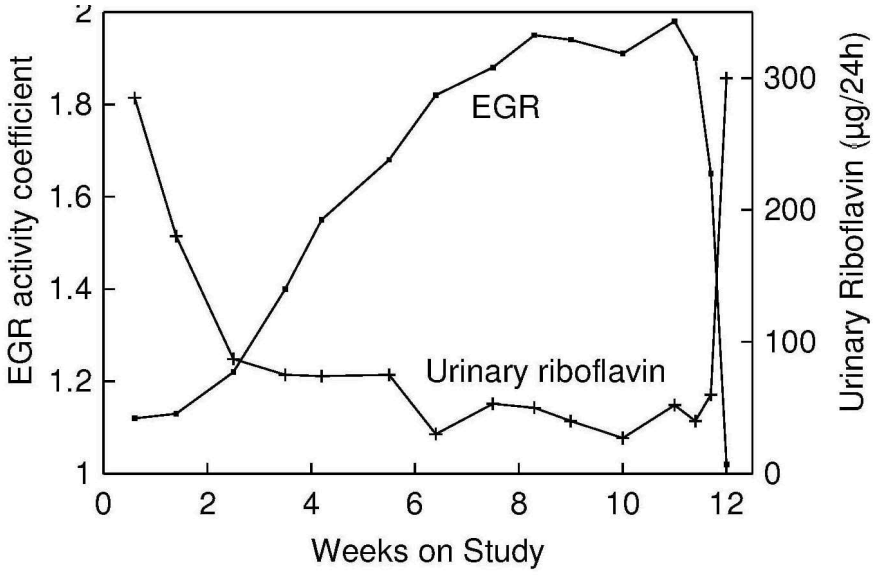


Figure 20b.2: Relationship during depletion / repletion of riboflavin intake to urinary riboflavin excretion and erythrocyte glutathione reductase activity (EGR) coefficients. The depletion phase (wk 2–10) involved riboflavin intakes of 0.07mg/d. Redrawn from Tillotson and Baker (1972).

Nevertheless, the increase in the EGR AC does not continue indefinitely; riboflavin intakes below 0.5mg/day do not produce any further increases in EGR AC (Sterner and Price, 1973). Consequently, the extent to which the EGR AC is elevated does not necessarily indicate the degree of riboflavin deficiency. Hence, it is not surprising that consistent correlations between EGR AC values and clinical signs of riboflavin deficiency have not always been observed (Bates et al., 1981).

Measurement of EGR AC has been used to monitor riboflavin status only in a few population-based studies such as the UK National Dietary and Nutrition Surveys Rolling Programme which includes all age and sex population groups (Bates et al., 2016), Irish National Adult Nutrition Survey (Kehoe et al., 2018) and the older national surveys from Germany(Heseker et al., 1992) and the Netherlands (Löwik et al., 1994). However, the assay has been used more frequently in cohort studies of UK older adults (Bailey et al., 1997), Irish adults (Jungert et al., 2020) and older people (Madigan et al., 1998; Moore et al., 2019), Canadian women of reproductive age (Aljaadi et al., 2019) and older adults (Whitfield et al., 2019), Spanish adults (Mataix et al., 2003; García-Minguillán et al., 2014) as well as French adults (Hercberg et al., 1994). Riboflavin status of different population groups assessed by EGR AC has been reported in studies conducted in developing countries including some in Africa(Bates et al., 1981; 1994; Ajayi, 1984; 1985; Ajayi and James, 1984), Central America (Boisvert et al., 1993a), and Asia (Thurnham et al., 1982; Blanck et al., 2002; Whitfield et al., 2015; Aljaadi et al., 2019).

Comparison of the prevalence of deficient riboflavin status among these studies is often difficult because the method of EGR AC assessment has not been standardized and as shown in a systematic evaluation of the existing methods, variations in the laboratory protocols applied in various studies could produce significantly different results (Hill et al., 2009). Moreover, the cutoff values used to denote suboptimal status and deficiency vary considerably from one study to another. However, EGR AC results from different populations generated by an identical protocol in one laboratory and by using the same cutoff showed very comparable data for the prevalence of riboflavin deficiency in Canadian and Irish older adults (26% vs 25.5%;Whitfield et al., 2019; Jungert et al., 2020) and in Canadian women of reproductive age and younger Irish adults (40% vs 52%; Aljaadi et al., 2019; Jungert et al., 2020) whereas 80% and 71% of the investigated Malaysian and Cambodian women of reproductive age, respectively were reported to be riboflavin deficient (Aljaadi et al., 2019; Whitfield et al., 2015).

In some of these surveys, significant inverse relationships between the EGR AC and riboflavin intakes have been observed (Gregory et al., 1990; Hercberg et al., 1994; Bailey et al., 1997; Bates et al., 1999; Hoey et al., 2007); results from the UK survey of British adults are given in Table 20b.1.

Table 20b.1: Correlation coefficients between the erythrocyte glutathione reductase activity coefficient and dietary riboflavin intakes from Gregory et al., 1990 .

	Erythrocyte glutathione reductase activity coefficient	
Dietary intake	**Men**	**Women**
Total riboflavin (incl. supplements)	−0.13 (p < 0.01)	−0.24 (p < 0.01)
Riboflavin from food sources	−0.23 (p < 0.01)	−0.31 (p < 0.01)

In the UK study of Norwich elderly (Bailey et al., 1997), initial EGR AC values for both males and females were significantly correlated with those measured 2y later (Table 20b.2), suggesting that EGR AC may be a reliable measure of long-term biochemical riboflavin status of individuals. This finding is consistent with earlier studies (Rutishauser et al., 1979).

There have been several attempts to establish a relationship in humans between a specific physiological function and EGR AC values, but the results have been inconclusive. Some reports of effects on work performance, neurovascular coordination, and iron handling with EGR AC values ≥ 1.7 have been reported (Prasad et al., 1990; Fairweather-Tait et al., 1992; Bates et al., 1994).

Table 20b.2: The correlation of EGR activity coefficients measured on 99 subjects initially and after 2y. Significant correlation does i between initial and follow-up mean values. Data from Bailey et al., British Journal of Nutrition 77: 225–242, 1997.

Elderly subjects	n	r	Significance
Males	37	0.411	< 0.02
Females	62	0.359	< 0.01

In some supplementation studies of subjects with subclinical riboflavin deficiency, no improvement in physical performance or endurance has been reported (Prasad et al., 1990; Winters et al., 1992; van der Beek et al., 1994). However, a supplemental riboflavin dose of 4mg/day given for 8wks appeared to have a beneficial effect on hematologic status in UK young women with a low riboflavin status (EGR AC > 1.65) at baseline (Powers et al., 2011). Furthermore, there is evidence that a low dose riboflavin supplementation can lower an elevated blood pressure, specifically in adults who are genetically predisposed to develop hypertension (Horigan et al., 2010; Wilson et al., 2013; see Section 20b.1.2 on deficiency of riboflavin in humans).

Some epidemiological studies also have linked EGR AC values with various health effects. In a study of adolescent Bhutanese refugees living in Nepal, for example, in whom the prevalence of angular stomatitis was 27%, those with angular stomatitis had significantly higher EGR AC values than those without (i.e., 2.2 ± 0.4 vs. 2.0 ± 0.3; p = 0.02). The adjusted odds ratio for angular stomatitis and low riboflavin status was 5.1 (95% CI: 1.55, 16.5) (Blanck et al., 2002).

In addition, a large cross-sectional study of Irish older adults showed that riboflavin deficiency (EGR AC > 1.46) was independently associated with an increased risk of depression after adjustment for covariates (OR 1.56, 95% CI: 1.10, 2.00, p=0.012) (Moore et al., 2019).

Factors affecting EGR AC values

FAD concentrations used in the assay to stimulate EGR can affect the EGR AC values obtained. Concentrations of FAD > 5µmol result in lower normal ranges of EGR ACs, in comparison with FAD concentrations ranging from 1–3µmol (Garry and Owen, 1976; Rutishauser et al., 1979; Hill et al., 2009).

The length of the pre-incubation of reagents with EGR enzyme appears also to be critical. A short pre-incubation period could underestimate EGR AC values because less time is available for FAD to bind to the enzyme in the hemolysate (Thurnham and Rathakette, 1982). Hill et al. (2009) found that 30min would be the optimal length of the pre-incubation.

Age of erythrocytes may also influence the EGR activity (Powers and Thurnham, 1981) as concentrations are declining with the aging of the cells.

Age of the subjects may affect the EGR activity. In some early studies a trend toward higher EGR ACs with increasing age, irrespective of sex, has been observed (Garry et al., 1982; Wright et al., 1995). However, more recent investigations including the UK population-based survey, have reported consistently higher EGR AC values in younger compared with older adults (Bates et al., 2016; García-Minguillán et al., 2014; Jungert et al., 2020). It is unknown whether this difference in riboflavin status with age

can be explained only by variations in dietary riboflavin intake.

Genetic disturbances, including glucose-6-phosphate dehydrogenase deficiency and heterozygous β-thalassemia, are associated with disorders in erythrocyte flavin metabolism (Prentice et al., 1981; Anderson et al., 1987; 1993), which can result in misleading EGR AC test results. For example, in cases of glucose-6-phosphate dehydrogenase deficiency, there is increased avidity of the EGR for FAD, resulting in EGR AC values within the normal range, even in the presence of clinical signs of riboflavin deficiency (Thurnham, 1972). Glucose-6-phosphate dehydrogenase deficiency is one of the most common enzyme defects in humans and it is highly prevalent among populations in malaria-endemic areas of sub-Saharan Africa and the Middle East where approximately up to 7.5% of people are affected (Nkhoma et al., 2009). It is estimated that glucose-6-phosphate dehydrogenase deficiency occurs in around 10% of Americans of African descent (Frischer et al., 1973). In contrast, in heterozygous β-thalassemia, there is an inherited slow red-cell metabolism of riboflavin to FMN and FAD and a high stimulation of the erythrocyte glutathionine reductase by extraneous FAD (Anderson et al., 1993).

Pyridoxine deficiency also interferes with the EGR AC test, resulting in a decreased erythrocyte glutathione reductase activity but no change in the activity coefficient, probably arising from a decrease in apoenzyme. No comparable effects have been observed for other vitamin deficiencies such as thiamin and vitamin C.

Disease states, including iron-deficiency anemia (Ramachandran and Iyer, 1974), severe uremia, cirrhosis of the liver, and hypothyroidism lead to increased erythrocyte glutathione reductase activity.

Conditions of negative nitrogen balance lead to a fall in EGR AC values. In pregnant women in The Gambia, for example, EGR AC values fell in association with a decline in body weight during the rainy season, despite evidence that malnutrition had actually increased (Bates et al., 1981). Similar findings have been reported in preschool children with upper respiratory tract infections and measles (Bamji et al., 1987).

Interpretive criteria

There is still uncertainty regarding the most appropriate threshold value indicative of normal riboflavin status. An EGR AC value of < 1.3 is generally considered to represent saturation of the tissues with riboflavin. This value was derived from the riboflavin depletion-repletion study in young men (Figure 20b.2) as well as from the supplementation study in elderly people with riboflavin deficiency (Boisvert et al., 1993b) where urinary excretion of riboflavin started to increase sharply (a point reflecting body saturation) at EGR AC below 1.3. A systematic review including 18 supplementation studies concluded that a cut-off of 1.3 should be considered as the “upper limit of a normal range” (Hoey et al., 2009). Sadowski 1992), however, used an upper limit of 1.34 in a study of apparently healthy elderly. This value was based on the mean plus 2 SD of the EGR AC value derived from a large sample of Boston elderly aged ≥ 60y. Cutoff EGR AC values of 1.3–1.4 for suboptimal and > 1.4 for deficiency states are often used (Wilson et al., 2013; Whitfield et al., 2015; Aljaadi et al., 2019; Jungert et al., 2020); these cutoffs appear to be independent of sex. Generally, EGR AC values in the range from 1.5 to 2.5 were found in population groups consuming very low intakes of riboflavin, wih evidence of clinical riboflavin deficiency (Bates et al., 1981; Thurnham et al., 1982; Lo, 1985). In a study of adolescent Bhutanese refugees, Blanck et al. (2002) used an EGR AC cutoff of > 1.7, and based on this value, 86% had low riboflavin status.

Considering that various methods for EGR AC assessment generate slightly different results (Hill et al., 2009), it is reasonable to expect that the cutoff value would be dependent on the analytical method used; thus, reference ranges developed in the local laboratories should be considered.

Table 20b.3: Median and 97.5thpercentile values for the activity coefficient (AC) of erythrocyte glutathione reductase (EGR) in UK ad (2016).

	EGR activity coefficient	
	19–64y	≥ 65y
Men Median	1.31	1.26
97.5th percentile	1.99	1.61

	EGR activity coefficient	
	19–64y	≥ 65y
Women Median	1.33	1.28
97.5th percentile	1.94	1.71

Information on the EGR activity coefficients from the UK National Dietary Nutrition Survey Rolling Programme, which includes data for all age groups split by sex is available (Bates et al, 2016). The median and upper 2.5thpercentile of the adult population by age and sex are presented in Table 20b.3

Measurement of erythrocyte glutathionine reductase activity

Only small samples of blood are required for the EGR assay, and fasting samples are not necessary (Komindr and Nichoalds, 1980). Either EDTA or heparin can be used as an anticoagulant. Erythrocytes must be washed and lysed, and the assay must be performed immediately after blood samples are drawn because FAD is a labile compound (Shenkin and Roberts, 2016). Alternatively, the hemolyzed samples can be kept frozen for over a year at –70°C without loss of EGR activity.

Generally, glutathione reductase activity is measured using an enzyme-coupled kinetic assay, although some colorimetric methods are also available (Sauberlich, 1984). An automated method involving the use of a centrifugal analyzer has been developed (Mak and Swaminathan, 1988; La Rue et al., 1997). An alternative procedure, using well plates and a plate reader, has also been described (Sauberlich, 1999). Very small volumes of whole blood instead of erythrocytes can be used for the measurement of glutathionine reductase activity, but the method, although simpler, is less sensitive. Care must be taken to ensure that the protocols for the assay are carefully specified and followed so that results can be compared among studies. As noted previously, the EGR assay is invalid for persons with glucose-6- phosphate dehydrogenase deficiency.

20b.2.2 Urinary riboflavin excretion

Flavins are excreted in urine mainly in the form of riboflavin (60–70% of all urinary flavins) or other metabolites, such as 7-hydroxymethylriboflavin (7-α-hydroxyriboflavin) and lumiflavin (Chastain and McCormick, 1987). Very little riboflavin can be stored in the body, so urinary excretion reflects dietary intake after tissues become saturated. A strong positive linear correlation of urinary riboflavin with riboflavin intake above 1.4mg/day has been reported (r^2 = 0.9667, p < 0.01) (Guo et al., 2016). Urinary riboflavin excretion is considered a biomarker of short-term status.

Experimental balance studies indicate that urinary riboflavin excretionary rates increase slowly with increasing intakes until the point of tissue saturation, after which any further increase of riboflavin intake leads to a sharp elevation of the excretion rate, as shown in Figure 20b.3. Based on the weighted results of experimental studies (Brewer et al., 1946; Horwitt et al., 1950; Boisvert et al., 1993b; Guo et al., 2016), the European Food Safety Authority estimated that the change of the urinary excretion rate of riboflavin corresponding to the point of tissue saturation could be achieved by a riboflavin intake of around 1.3mg/d (EFSA, 2017). Once intakes of 2.5mg/d are reached, excretion becomes approximately equal to the rate of absorption (Horwitt et al., 1950). At such high intakes, a significant proportion of the riboflavin intake is not absorbed.

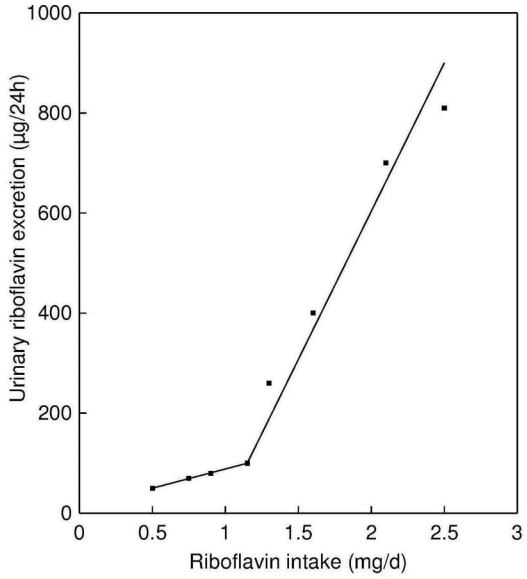


Figure 20b.3: Relationship of riboflavin intake to urinary excretion. Redrawn from Sauberlich (1999).

Urinary riboflavin was measured in several early population studies in North America: the Nutrition Canada survey, the US Ten State Nutrition Survey, and in NHANES I. In these surveys, casual urine samples were collected rather than the preferable 24h urine samples. As a result, riboflavin concentrations were expressed per gram of creatinine. For children, the rate of riboflavin excretion, when expressed per gram of creatinine is greater than for adults. The population-based VERA Study, conducted as part of the German National Consumption Study I, assessed urinary riboflavin in 24hour urine samples of 2,006 adults (Heseker et al., 1992)

Occasionally, a riboflavin load test is used to assess the degree to which the body is saturated with riboflavin. An oral dose of 5mg riboflavin is given, and again the timed excretion of riboflavin in the urine is measured, usually over a 4h period (Lossy et al., 1951; ICNND, 1963;Brun et al., 1990); the results should be compared with the level in the urine before the test load.

Factors affecting riboflavin excretion

Physical activity and sleep can decrease riboflavin excretion (Soares et al., 1993).

Negative nitrogen balance and infection induce a breakdown of tissue protein and, hence, increase urinary excretion of riboflavin (Tucker et al., 1960). Consequently, measurement of urinary riboflavin is not very useful in circumstances where intakes of protein are low or where chronic infection is common.

Drugs including some antibiotics and psychotropic drugs such as phenothiazines can increase the excretion of riboflavin (Goldsmith, 1975).

Oral contraceptive agents and pregnancy influence riboflavin excretion. Concentrations decrease when oral contraceptive agents are used, and during the third trimester of pregnancy, as tissue retention of the vitamin is increased in response to increased need. However, during the second trimester, excretion increases.

Within-subject variability can be large for 24h urinary riboflavin excretion. Coefficients of variation range from 13% to 25%, which are not reduced when results are expressed in terms of creatinine (van Dokkum et al., 1990).

Interpretive criteria

Interpretive criteria for urinary riboflavin excretion for adults and children are shown in Table 20b.4.

Table 20b.4: Interpretive criteria for the urinary excretion of riboflavin. tri: trimester. From (Sauberlich HE ,1999).

Subjects	Less than acceptable (at risk)		Acceptable (
	Deficient (high risk)	Suboptimal (med. risk)	
Children	(µg/g creatinine)		
1–3y	< 150	150–499	≥ 500
4–6y	< 100	100–299	≥ 300
7–9y	< 85	85–269	≥ 270
10–15y	< 70	70–199	≥ 200
Adults	< 27	27–79	≥ 80
Preg. 1st tri.	< 27	27–65	66–129
Preg. 2nd tri.	< 23	23–54	55–109
Preg. 3rd tri.	< 21	21–49	50–99
Adults	Other interpretive guidelines		
µg/24h	< 40	40–119	≥ 120
µg/6h	< 10	10–29	≥ 30
µg in 4h after ribofl. load	< 1000	1000–1399	≥ 1400

For adults, when dietary intakes are adequate, ≥ 120µg of riboflavin per day should be excreted. After a loading dose of 5mg riboflavin, then ≥ 1400µg of riboflavin should be excreted in a 4h period under conditions of adequate riboflavin intake. In a deficiency state, < 1000µg of riboflavin will be excreted during the same time period. Interpretive criteria for riboflavin excretion in casual urine samples expressed as µg/g creatinine for children, adults, and pregnant women are also given in Table 20b.4.

Measurement of urinary riboflavin

In the past, fluorometric and microbiological assays have generally been used to determine riboflavin concentrations in urine. The fluorometric method measures the fluorescence of the flavins directly or converts the flavins to lumiflavin and then determines the fluorescence. The microbiological methods use *Ochromonas danica* or the protozoan *Tetrahymena pyriformis*.

Measurement of urinary riboflavin using HPLC with fluorometry (Gatautis and Naito, 1981) or competitive protein-binding assays (Tillotson and Bashor, 1980) is also possible. The protein-binding assays are rapid, sensitive, and require no treatment of the urine sample prior to analysis. Discrepancies among the analytical methods have been documented. Results for the HPLC method with fluorometry tend to be lower than those with the older fluorometric method alone because, in the HPLC method, riboflavin is separated from other flavins (Smith, 1980).

20b.2.3 Riboflavin, FMN and FAD in blood

Plasma or serum concentration of riboflavin and its derivatives, FMN and FAD have been used in a few studies for assessment of riboflavin status. The results of these investigations showed that supplementation with low doses of riboflavin increased plasma free riboflavin (Bessey et al., 1956; Hustad et al., 2002; Guo et al., 2016) and FMN (Bessey et al., 1956; Hustad et al., 2002) whereas no detectable or very modest changes were reported for plasma FAD concentrations (Bessey et al., 1956; Zempleni et al., 1996; Hustad et al., 2002).

The concentration of riboflavin forms in erythrocytes could be considered as marker of long-term riboflavin intake. Erythrocyte concentration of riboflavin (mainly FMN and FAD) was found to be significantly lower in individuals with clinical signs of riboflavin deficiency compared with healthy people (Bamji, 1969). Supplementation with a low dose of riboflavin for 12wks increased erythrocyte FMN by 87% whereas more modest but still significant response was reported for erythrocyte FAD (Hustad et al., 2002). A good relationship was shown between EGR AC and erythrocyte riboflavin (Bates et al., 1999) and erythrocyte FMN and FAD (Hustad et al., 2002) indicating that these measurements have potential and warrant further exploration.

Fluorometric and microbiological assays have been used to determine riboflavin in plasma / serum and erythrocytes. Hustad et al. (1999) have developed a sensitive and robust method, based on capillary electrophoresis and laser-induced fluorescence detection, for quantifying the low physiological concentrations of FMN and FAD, as well as riboflavin, in human plasma; later the method has been developed further for assessment of riboflavin derivatives in erythrocytes (Hustad et al., 2002).

20b.2.4 pyridoxamine phosphate oxidase (PPO) activity and activity coefficient

The measurement of erythrocyte PPO enzyme activity and its activity coefficient (PPO AC; assessed by the ratio of PPO activity before and after *in vitro* activation with the prosthetic group FMN) have been suggested as alternatives to EGR AC for assessment of riboflavin status. This option could be very useful for diagnosis of riboflavin deficiency in populations with a high prevalence of glucose-6-phosphate dehydrogenase deficiency since the EGR AC assay produces erroneous results in individuals with this genetic defect (Thurnham, 1972; Prentice et al., 1981).

A fluorometric assay for measurement of PPO activity in red blood cell lysates has been developed (Bates and Powers, 1985) and subsequently optimized and adapted for assessment of PPO AC (Mushtaq et al., 2009). The results generated by this methodology demonstrated that both PPO and PPO AC are responsive to riboflavin supplementation and PPO showed a strong dose-response effect. In addition, in people free from glucose-6-phosphate dehydrogenase deficiency, PPO and PPO AC strongly correlated with the well-established biomarker of riboflavin status, EGR AC (Mushtaq et al., 2009). Therefore, both PPO and PPOAC are promising biomarkers, however, criteria to assess riboflavin adequacy based on these biomarkers have not been developed yet.

This page titled 19.2: Biomarkers of riboflavin status (20b.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

20: NIACIN (CHAPTER 20C)

Abstract

The term "niacin" refers to three compounds: nicotinamide, nicotinic acid, and nicotinamide ribose. They are precursors of nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP), both of which function in a variety of oxidation and reduction reactions and are associated with both catabolic and anabolic processes. Pellagra, a disease of the "4Ds" (diarrhea, dermatitis, dementia, and death) is the syndrome of niacin deficiency. Although rare, outbreaks of pellagra still occur when diets are based on maize (without nixtamalization), a cereal low in bioavailable niacin. Secondary niacin deficiency may also develop in diseases that interfere with absorption of niacin or tryptophan (a precursor of niacin) (e.g., chronic colitis, liver cirrhosis) or reduce the conversion of tryptophan to niacin (e.g., chronic alcoholism, prolonged isoniazid treatment for tuberculosis). Rich food sources of niacin include meat, poultry, and fish, followed by dairy products and enriched or fortified cereals. Coffee is also a good source of niacin. Absorption of niacin from food ranges from about 23% to 70% and is lowest from cereals and highest from animal products. Certain food preparation and processing practices may influence the content and bioavailability of niacin in food. Intakes and requirements for niacin are expressed as niacin equivalents (NEs) to consider the contribution of tryptophan to niacin (i.e., 60mg tryptophan yields 1mg niacin). Dietary reference values for niacin are set as NE per day, or alternatively expressed as mg NE/MJ or mg NE/1000 kcal based on the relationship between the requirements for niacin and energy. Niacin toxicity may occur when large pharmacological doses of niacin are taken to prevent or treat metabolic diseases such as dyslipidemia. Tolerable Upper Intake Levels (ULs) are based on dermal vasodilative flushing.

At present, the best biomarker of niacin status is the measurement of two urinary metabolites: N'-methylnicotinamide and N'-methyl-2-pyridine-5-carboxamide (2-pyridone) using HPLC. Sampling strategies employed include 24h urines with results expressed as concentrations of individual metabolites (mg/d), or random spot urines when the ratio (2-pyridone to N'-methylnicotinamide), or concentrations relative to creatinine (mg/g creatinine) are used. Interpretive criteria for adults are available for urinary excretion of these niacin metabolites expressed relative to creatinine and as the ratio (2-pyridone / N'-methylnicotinamide). WHO has developed provisional criteria for the severity of public health problem of niacin deficiency. Concentrations of the co-enzyme NAD in erythrocytes (but not whole blood) or the ratio of erythrocyte NAD to NADP (called niacin index) are said to be sensitive to short-term changes in niacin intake, even in the absence of clinical signs of deficiency. Hence, they may serve as a sensitive and reliable functional biomarker for risk of niacin deficiency. With the development of a rapid LC-MS/MS method for their analysis, further investigation of erythrocyte NAD and NADP concentrations as biomarkers of niacin status is warranted.

20.1: Niacin - Introduction (20c.1)

20.2: Indices of niacin status (20c.2)

This page titled 20: Niacin (Chapter 20c) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

20.1: Niacin - Introduction (20c.1)

Niacin (also known as "vitamin B3") is the term used to describe a group of compounds with the biological activity of the vitamin. They include nicotinic acid (also known as niacin) composed of a pyrimidine ring bound to a carboxylic group (pyridine-3-carboxylic acid), nicotinamide (also known as niacinamide) composed of a pyrimidine ring bound to a carboxamide group (pyridine-3-carboxamide), and nicotinamide riboside. The third compound, nicotinamide riboside has been identified more recently and is a pyridine-nucleoside form of vitamin B3 consisting of nicotinamide with a beta-D-ribofuranosyl moiety at the 1-position (Mehmel et al., 2020). All three compounds have similar but not identical properties and are precursors of nicotinamide adenine dinucleotide (NAD) and its phosphorylated derivative, nicotinamide adenine dinucleotide phosphate (NADP). The structures of nicotinic acid, nicotinamide, and nicotinamide riboside are shown in Figure 20c.1.

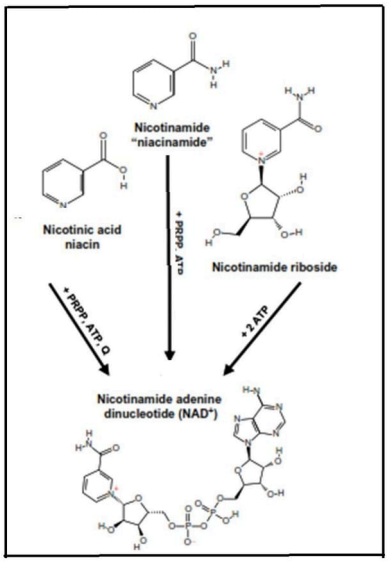


Figure 20c.1. Vitamin B3 molecules (nicotinic acid, nicotinamide, and nicotinamide riboside), dietary precursors that support the formation of nicotinamide adenine dinucleotide. Redrawn from Kirkland & Meyer-Ficca (2018).

20c.1.1 Functions of niacin

Most of the niacin in food is present as NAD and NADP; very little exists as free forms of niacin (Chungchunlam & Moughan, 2023). Prior to absorption, these two components must first be hydrolyzed to free nicotinamide by phosphatases and NAD glycohydrolases in the small intestine, the major site of absorption. At low intakes, absorption into the intestinal cell is by sodium-dependent, carrier-mediated diffusion, but at higher intakes, passive diffusion predominates (Hrubša et al., 2022). Once absorbed, free nicotinamide is transported to all tissues for the synthesis of the two coenzymes NAD and NADP within the cells. In general, concentrations of intracellular NAD are higher than NADP. All tissues of the body can synthesize NAD and NADP, although concentrations are greatest in the liver where some storage may occur. Some niacin can be synthesized by colonic bacteria in the large intestine (Yoshii et al., 2019). NAD exists in two forms: NAD+ (oxidized) and NADH (reduced), while NADP exists as NADP+ (oxidized) and NADPH (reduced).

NAD+ can be synthesized within the cell from nicotinic acid, nicotinamide, nicotinamide riboside, and tryptophan by three pathways, all of which require the enzyme nicotinamide mononucleotide phosphoribosyltransferase (NMNAT). The three pathways include the conversion of tryptophan to NAD via the *de novo* NAD synthesis pathway, the Preiss-Handler pathway from nicotinic acid, comprising three individual metabolic conversions, and the salvage pathway involving the NAD+ precursors nicotinamide or nicotinamide riboside. Of the three, the salvage pathway is the dominant one; for more details on these synthetic pathways see Kirkland & Meyer-Ficca (2018).

NAD+ primarily participates in catabolic reactions and is an essential substrate for the activity of three classes of NAD+ -consuming enzymes: (i) sirtuins (SIRTs), (ii) poly(ADP-ribose) polymerases (PARPs), and (iii) cyclic ADP-ribose (cADPr) synthases. These NAD+-consuming enzymes cleave NAD+, resulting in an ongoing loss of NAD+ during which nicotinamide is produced as a reaction by-product. Because loss of NAD+ cannot be compensated by intake of NAD from the diet, nicotinamide is recycled into NAD+ via the salvage pathway. This pathway involves the activity of NAMPT and nicotinamide mononucleotide adenylyltransferase (NMNAT 1-3). Hence, activity of these enzymes has a larger effect on NAD concentrations than nicotinamide or dietary intakes of niacin. For more details on the action of these enzymes see Xiao et al. (2018) and Penberthy & Kirkland (2020).

NAD+ can be converted to NAD+phosphate (NADP+) by the activity of NAD+ kinase. NAD(H) can also be utilized as a substrate during which NADPH is produced instead of NADP+, a process involving yet another NAD+-consuming process. During metabolism both NAD+ and the phosphorylated form NAD(P)+ undergo redox reactions and generate two redox couples NAD+ / NAD(H) and NAD(P)+ / NAD(P)H. The NAD+ / NAD(H) redox couple participates in about 400 reactions, predominately in catabolic metabolism, and regulates cellular energy metabolism, that is glycolysis and mitochondrial oxidative phosphorylation. By contrast, the NADP+ / NAD(P)H pool participates in about 30 reactions, mainly in anabolic metabolism, and supports the biosynthesis of fatty acids, nucleic acids and cholesterol / steroids. It is also involved in maintaining redox balance. For a review of the catabolic processes and anabolic processes involving NAD+ and NAD(P)+, see Hrubsa et al. (2022) and Kirkland & Meyer-Ficca (2018) for more details.

Loss of cellular redox homeostasis of NAD+ / NAD(H) has been linked to a variety of pathological conditions, such as cardiovascular diseases, neurogenerative diseases, cancer, and aging (Xiao et al., 2018). More research is required to better understand the cellular functions of NAD(H) and NADP(H) and how loss of redox homeostasis affects energy metabolism.

In addition to its redox roles, NAD+ is also an essential cofactor for important non-redox signaling pathways, thus regulating biological functions, including gene expression, cell cycle progression, DNA repair and cell death. For more details on the role of NAD in the central nervous system, see Gasperi et al. (2019).

The major pathway of catabolism of nicotinic acid and nicotinamide is by methylation in the liver to N'-methylnicotinamide (NMN) and subsequent oxidation to N-methyl-2-pyridone-carboxamide (2-pyridone) and N'-methyl-4-pyridone-5-carboxamide (4-pyridone). Under normal conditions, 20-35% and 45-60% of niacin is excreted as N'-methylnicotinamide (NMN) and 2-pyridone respectively, although the amount varies depending on niacin and tryptophan intake (EFSA, 2014).

A small percentage of the essential amino acid tryptophan is converted directly to nicotinamide adenine dinucleotide (NAD) via a two-step pathway (Figure 20c.2), primarily in the liver. The conversion rate is dependent on the status of tryptophan rather than niacin, and is controlled by the activities of several enzymes; for more details see Fukowatari & Shibata (2013). Low intakes of tryptophan, interference with tryptophan transport, as occurs in Hartnup's disease, and conditions associated with increased tryptophan metabolism (e.g., inflammatory bowel diseases) (Nikolaus et al., 2017), all reduce the conversion of tryptophan to niacin, resulting in an increased need for preformed niacin. Pyridoxal 5-phosphate (a vitamin B6 derived coenzyme), iron, riboflavin, ascorbic acid, and glutamine are involved as cofactors in the conversion of tryptophan to NAD (Penberthy & Kirkland, 2020).

Physiological amounts of both nicotinic acid and nicotinamide prevent the classical signs of niacin deficiency (see section below). However, certain niacin compounds also have pharmacological properties. At supraphysiological doses, nicotinic acid is known to decrease total cholesterol, LDL cholesterol, and triglycerides and increase high-density lipoprotein cholesterol (HDL-C) (Minto et al., 2017). Chronic supplementation with nicotinamide ribose may also have some positive benefits such as reducing blood pressure and arterial stiffness, and warrants further study (Martens et al., 2018). Recent research has also implicated niacin as a therapeutic neuroprotective agent (Gasperi et al., 2019), although more studies are needed to confirm the role of niacin for alleviating neurological impairment (Wuerch et al., 2023). In addition, decline of NAD+ levels during the aging process has led to investigations on the role of niacin in aging-related processes. Results suggest that niacin supplements, as the major nutritional precursor of NAD, may provide antiaging properties (Imai et al., 2014).

20c.1.2 Deficiency of niacin in humans

Pellagra is the characteristic syndrome of severe niacin deficiency and its dietary precursor tryptophan. Pellagra is known as the disease of the "4Ds" (diarrhea, dermatitis, and dementia, culminating in death). Early symptoms include diarrhea resulting from intestinal inflammation, accompanied by anorexia, nausea, and abdominal pain. Later, dermatitis develops in sun-exposed skin

such as dorsal parts of hands, feet, cheeks, forehead, and sun-exposed regions of the neck (WHO, 2000). These signs arise from deficits in poly (ADP-ribose) polymerase activity in response to UV radiation-induced DNA damage. The third D refers to dementia and presents with hallucinations and delusions, symptoms said to be caused by impaired formation of cADP-ribose and nicotinic acid ADP, which result in alterations in neural calcium signaling (Meyer-Ficca & Kirkland, 2016).

Migraine, a neurological condition related to brain energy deficiency, has recently been associated with intakes of dietary niacin in a population-based study of US adults (>20y) participating in NHANES 1999-2004 (Liu et al., 2022). An L-shaped relationship between dietary niacin intake and migraine was reported, with the risk decreasing with increasing dietary niacin consumption in those with a dietary niacin intake of <21.0mg/d. However, there was no association between dietary niacin consumption and migraine when the daily niacin intake was >21.0mg/d, indicating that the risk of migraine no longer decreases with increasing dietary niacin intake (Liu et al., 2022). A few case reports have also suggested that niacin may be effective as an adjuvant treatment for acute migraine, although the underlying mechanism is uncertain.

Pellagra is often a problem in countries where maize and millet jowar (*Sorghum vulgare*) are dietary staples. The bioavailability of niacin in maize is poor (only about 30%) because most of the niacin is present in two forms: niacytin and niacinogenes. Niacytin consists of nicotinic acid esterified to polysaccharides, whereas niacinogenes are bound to polypeptides and glycopeptides; both forms are unavailable for absorption by intestinal enzymes. Maize is also particularly low in tryptophan, a precursor of NAD (WHO, 2000). In contrast, millet jowar contains suitable amounts of tryptophan, but excessive amounts of leucine which interfere with the conversion of tryptophan to niacin. The mechanism is uncertain. Excess leucine may disrupt the availability or utilization of riboflavin and vitamin B6, micronutrients required in the synthesis of niacin from tryptophan (Kirkland & Meyer-Ficca, 2018).

Niacin fortification of cereals, notably maize, has been associated with widespread reductions in the prevalence of pellagra (Vilijoen et al., 2022). Nevertheless, outbreaks of pellagra still occur where cereals are not fortified and diets are restricted. Such outbreaks have been reported in Malawi among Mozambican refugees (Malfait et al., 1993), and more recently, among adults in the Kasese Catchment Area, Dowa, Malawi (Matapandeu et al., 2017).

Secondary pellagra can develop from certain diseases or conditions that interfere with the absorption of niacin or tryptophan or reduce the conversion of tryptophan to niacin. Diseases that interfere with absorption may include chronic diarrhea, chronic colitis, cirrhosis of the liver, and tuberculosis of the gastrointestinal tract (Prabhu et al., 2021). Chronic alcoholism causes a reduction in the conversion of tryptophan to niacin, arising from inhibition of tryptophan 2,3-dioxygenase, a rate-limiting enzyme of the hepatic kynurenine pathway of tryptophan degradation and niacin synthesis (Prabhu et al., 2021). Even in countries with niacin fortification of wheat flour such as the United States, alcohol-associated pellagra cases have been reported. Other diseases in which the conversion of tryptophan to niacin is reduced include the genetic disorder Hartnup disease when there is a deficit in tryptophan transport (Kirkland & Meyer-Ficca, 2018) and carcinoid syndrome in which tryptophan is preferentially oxidized to 5-hydroxytryptophan and serotonin (Gade et al., 2020).

Prolonged treatment of tuberculosis with the drug isoniazid also reduces the conversion of tryptophan to niacin by competing with pyridoxal 5'-phosphate (a vitamin B6 derived coenzyme required in the tryptophan-to-niacin pathway). In Malawi, an increased risk of pellagra has been associated with mass scale-up of isoniazid use for the preventive treatment of tuberculosis for people living with HIV. Continuous isoniazid preventive treatment for tuberculosis and the annual period of food scarcity were reported to increase the risk in this matched case-control study (Nabity et al., 2022). Such findings indicate that co-administration of niacin-containing multi-B vitamins with isoniazid as a pellagra prevention, warrants exploration.

20c.1.3 Food sources and dietary intakes

Most of the niacin in food is present as a component of NAD or NADP as noted earlier; very little exists as free forms of niacin, with the exception of liver and beans. Rich food sources of niacin include meat (especially liver), poultry, and fish followed by dairy products, oilseeds, some cereals (especially when enriched or fortified with niacin), legumes (including peanuts) and baker's yeast (Hrubša et al., 2022). In milk, 40% of the niacin is present as nicotinamide riboside and 60% as nicotinamide (Bieganowski & Brenner, 2004). Animal products release nicotinamide from its nucleotide forms during food processing and digestion, whereas plant products largely deliver nicotinic acid. Tea, coffee, and cocoa beverages also contains an appreciable amount of niacin.

Absorption of niacin from food ranges from about 23% to 70% and is lowest from cereals and highest from animal products (EFSA, 2014). Certain food preparation and processing methods can influence the content and bioavailability of niacin in food. Milling of cereals reduces their niacin content. Treatment of cereals with alkali (e.g., lime water), baking with alkaline baking

powders, and roasting whole grain maize, all increase niacin bioavailability by releasing the bound forms of niacin. The thermo-alkaline processing of maize kernels termed "nixtamalization" and practiced in Latin America, is responsible, at least in part for the very low prevalence of pellagra in the region (Hrubša et al., 2022). Roasting green coffee also increases its nicotinic acid content, by removing the methyl group from trigonelline (N'-methylnicotinic acid) (Hrubša et al., 2022). The coffee cultivar, degree of roasting, and brewing techniques all influence the amount of nicotinic acid in a cup of coffee. Although resistant to high temperatures, significant amounts of water-soluble niacin can be lost in cooking water, if discarded. More details of the content and bioavailability of niacin in both animal and plant-based foods are available in Chungchunlam & Maughan (2023).

In industrialized countries, meat and meat products often contribute as much as one third of the total intake of niacin of adults, closely followed by enriched or fortified cereals, whole-grain breads and bread products, and fortified ready-to-eat cereals; vegetables and milk + milk products contribute about one-tenth of the total niacin intake (Gregory et al., 1990; McLennan and Podger, 1995; IOM, 1998; EFSA, 2014).

Because niacin can be derived from the amino acid L-tryptophan, a precursor of niacin, intakes are usually expressed in terms of niacin equivalents (NE). About 60mg of tryptophan yields 1mg niacin following digestion and absorption. However, the conversion rate varies widely (30%) between individuals and tends to increase with increased tryptophan consumption. The most common sources of tryptophan in U.S diets are high protein foods such as fish, chicken, and milk, as well as bananas, chocolate, and peanuts (Richard et al., 2009). Conversion of niacin from tryptophan does not meet the need for niacin, of which about 50% is provided by dietary niacin sources. In Europe in 2014, the average daily intakes of total niacin from nine countries ranged from 27 to 53mg NE/day, with those for males slightly higher than females due mainly to the larger quantities of food consumed(EFSA, 2014). A summary table of the niacin intakes in these nine European countries is available in ESFA (2014).

In the United States, the Dietary Reference Intakes (DRIs) are set by the Institute of Medicine (IOM, 2000). The Estimated Average Requirements (EARs) are 12mg NE/day and 11mg NE/day, respectively for male and female adults. Corresponding levels for the US Recommended Dietary Allowances (RDAs) are 16mg NE/day for males and 14mg NE/day for females. Both male and female RDA intake levels for adults are similar to the Recommended Nutrient Intakes (RNIs) set by WHO/FAO (2004). In contrast, EFSA (2014) have set the Dietary Reference Values (DRVs) for niacin based on the relationship between niacin requirement and energy requirement. The Average Requirement (i.e., EAR) for adults (both men and women) is set at 1.3mg NE/MJ (about 5.5mg NE/1000kcal) and the Population Reference Intake (i.e. the RDA & RNI) corresponds to 1.3mg NE/MJ (about 6.6mg NE/1000kcal) assuming a coefficient of variation of 10%. These are the same levels set by the Scientific Committee for Food (SCF) in 1993. For details of the DRVs for other age and life-style groups set by these agencies, see relevant publications.

20c.1.4 Effects of high intakes of niacin

Niacin toxicity may occur when large pharmacological doses of different forms of niacin, especially nicotinic acid, are taken to prevent or treat some metabolic diseases, as noted earlier. For example, nicotinic acid is used in high doses (>1000mg/d) to treat hyperlipidemia but is sometimes accompanied by side effects. Of these, facial flushing caused by prostaglandin D2-mediated vasodilatation of small subcutaneous blood vessels, may occur, and has even been reported at doses of nicotinic acid below 50mg/d (MacKay et al., 2012).

Most agencies use dermal vasodilative flushing as the basis of their Tolerable Upper Intake Levels (ULs) for niacin compounds. The U.S. Food and Nutrition Board (IOM, 2000) have set the Tolerable Upper Intake Level (UL) for adults at 35mg/d for all forms of "niacin", whereas both Europe and the United Kingdom have set different ULs for free nicotinic acid and nicotinamide because nicotinamide does not produce vasodilative flushing but has no beneficial effects on lipid profiles (Minto et al., 2017). For free nicotinic acid and nicotinamide, the ULs for adults set by the European Food Safety Authority (ESFA) are 10mg/d and 900mg/d, respectively (EFSA, 2014), and 17mg/d and 500mg/d, respectively for the UK Safe Upper Levels (SULs)(Food Standard Agency, 2003). Differences in the methods used by these agencies to set the ULs may account in part for these discrepancies (Minto et al., 2017). There is some concern about whether the flushing effect is an appropriate basis for the UL for nicotinic acid (MacKay et al., 2012), and whether different ULs should be set for healthy and unhealthy persons (Minto et al., 2017). Note: the UL does not apply to persons taking high-dose niacin for treatment of dyslipidemia.

This page titled 20.1: Niacin - Introduction (20c.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

20.2: Indices of niacin status (20c.2)

More recent research has focused on the effects of supraphysiological doses of niacin compounds and their potential health benefits. Fortification of cereals with multiple micronutrients, including niacin, has been accompanied by a reduction in the prevalence of pellagra in low- and middle income countries and may account for the lack of research on biomarkers of status and function for niacin. The tests available today reflect only recent dietary intakes of niacin and therefore may not identify all persons at risk to deficiency.

Measurement of the urinary metabolite of N'-methylnicotinamide, in combination with one or more of the urinary pyridone turnover products, remains the biochemical test recommended by WHO (2000) to survey at risk populations to determine the extent and severity of deficiency of niacin and/or tryptophan. However, as simpler, and more rapid methods for measuring NAD and NADP in erythrocytes become available in the future, investigations of their use as indices of niacin status may increase. The current tests based on niacin metabolites in urine and erythrocytes are described below.

20c.2.1 Urinary excretion of niacin metabolites

Urine is the main route of excretion for niacin, with N'-methylnicotinamide and N'-methyl-2-pyridone-5-carboxamide (2-pyridone) being the major end products of niacin metabolism. These products are derived from either preformed niacin or niacin obtained from dietary tryptophan Figure 20c.2.

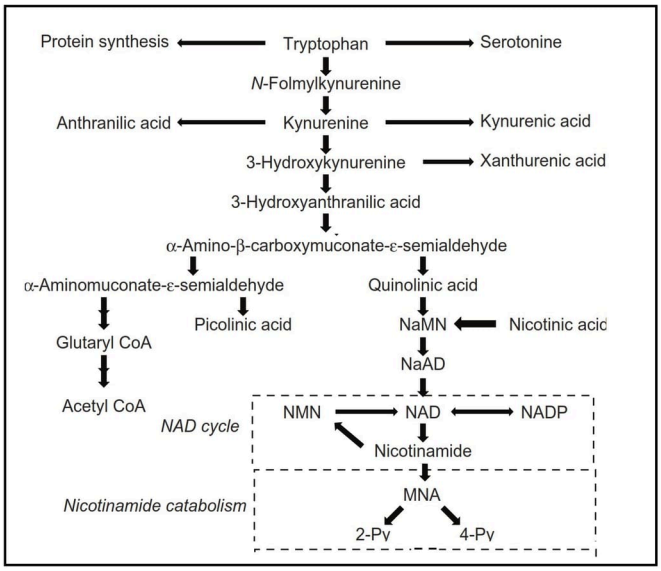


Figure 20c.2 The tryptophan - nicotinamide pathway. Note the pathway consists of the two parts: the first part is from tryptophan to quinolinic acid and the second is from quinolinic acid to 2-pyridone and 4-pyridone that includes the NAD cycle and nicotinamide catabolism. Abbreviations: NaMN: nicotinic acid mononucleotide; NMN: nicotinamide mononucleotide; MNA: N'-methylnicotinamide; 2-pyr: N'-methyl-2-pyridone-5-carboxamide; 4-pyr: N'-methyl-4-pyridone-3-carboxamide. From Fukuwatari and Shibata (2013).

Healthy adults generally excrete 20% to 35% of nicotinic acid as the N'-methylnicotinamide (NMN) form and 45%-60% as 2-pyridone (de Lange and Joubert, 1964), although this pattern does vary with the amount and form of niacin ingested and the niacin status of the subject.

The excretion of both N'-methylnicotinamide and N'-methyl-2-pyridone-5-carboxamide (2-pyridone) decreases in niacin deficiency, and their measurement in 24h urine samples is considered the most reliable and sensitive measures of the adequacy of intakes of niacin and tryptophan, a precursor of nicotinic acid. For example, in an experimental study of Jacob et al. (1989), during the depletion phase there was a significant fall in both 24h urinary N'-methylnicotinamide and 2-pyridone excretion after 35 days on a "low" intake of 6.1NE/d. Based on these findings, measurement of these two urinary metabolites appears reliable for assessing very low niacin intakes (i.e., 6.1NE/d). Of the two, however, N'-methylnicotinamide excretion is the easiest to measure, and is more sensitive to marginal intakes (i.e., 10.1NE/d) than urinary 2-pyridone (Jacob et al., 1989) (Table 20c.1).

Table 20c.1 Variation in the concentration of urinary metabolites during a niacin depletion / repletion study on healthy young men. T period. NMN, (N'-methylnicotinamide; 2-pyr, 2-pyridone; Data from Jacob et al., 1989.

Metabolic period	Niacin intake (NE/d)	n	Day
Stabilization	19.6	7	11–13
Depletion	6.1	3	47–49
Depletion	10.1	4	47–49
Repletion	19.2	6	62–64
High intake	25	3	77–79
High intake	32	3	77–79

In a later study in Angola of patients with clinical pellagra, low urinary excretion of the two urinary metabolites were shown to be a sensitive and specific indicator of clinical pellagra (Creeke et al., 2007).

Results of more recent observational studies are consistent with these earlier findings, and indicate that 24h levels of urinary niacin metabolites can be used as a biomarker of dietary intake of niacin. Significant linear correlations were observed between 24h urinary excretion of the sum of the three niacin metabolites (N'-methylnicotinamide, 2-pyridone, and N'-methyl-4-pyridone-3-carboxamide) and usual dietary niacin intake (expressed as niacin equivalents) in both apparently healthy college students (Tsuji et al., 2010) and school children (aged 10-12y) in Japan (Tsuji et al., 2011).

In the early experimental studies and the observational surveys conducted in Japan, the urinary metabolites were measured in 2h (de Lange and Joubert, 1964), or 24h collections (Jacob et al., 1989; Tsuji et al., 2010, 2011), and expressed as a concentration (mg/d, Table 20c.1) for the metabolites. However, the collection of timed urine samples is difficult in field studies. Instead, collection of random spot urine samples is a more practical sampling alternative, although few studies have investigated this approach.

In a pilot study, Creeke and Seal (2005) analyzed both N'-methylnicotinamide and 2-pyridone levels simultaneously in spot urine samples collected from two healthy subjects over four consecutive days. In the early morning of each day, an overnight fasting urine sample was collected. In addition, on day 4, non-fasting urine samples were collected at 3h intervals over 9h from each subject. Use of vitamin supplements by each subject was recorded. Their preliminary findings suggest that the use of fasting, early morning urine samples for assessing niacin status has potential: the impact of recent dietary intake is reduced and the samples are more indicative of longer-term niacin status. They also noted that expression of the metabolites relative to creatinine in the fasting spot urine samples yielded more stable results. Nevertheless, they caution use of this expression when applied to individuals with protein malnutrition and thus reduced creatinine output. In such circumstances, use of the urinary metabolite / creatinine ratio is inappropriate because the resultant ratio will be suggestive of a higher niacin status (Dillon et al., 1992).

In a later study in Angola, Creeke and co-workers (2007) used fasting spot urine samples to measure and compare the urinary metabolites of niacin — N'-methyl-2-pyridone-5-carboxamide (2-pyridone) and N'-methylnicotinamide — collected from both healthy individuals (n=2) and from patients diagnosed with pellagra (n=34). In this study, results were expressed as the concentration of the metabolites relative to creatinine to provide the most stable measurement of status. When expressed in this way, individuals with clinical pellagra had lower concentrations of the two urinary metabolites, which rose markedly following

treatment with a nicotinamide supplement. These findings indicate that the use of spot urine sample 2-pyridone and N'-methylnicotinamide concentrations, relative to creatinine, are a sensitive and specific measure of severe deficiency.

Based on earlier research, a ratio of 2-pyridone to N'-methylnicotinamide has been proposed as a convenient alternative index of niacin status, independent of age and creatinine excretion, and applicable to casual urine samples (de Lange and Joubert, 1964). However, the use of this ratio, has been challenged. Although useful for detecting severe niacin deficiency accompanied by pellagra, the ratio appears less useful in less severe deficiency states. For example, in the experimental depletion-repletion study in adult males (Jacob et al., 1989), the ratio of 2-pyridone / N'-methylnicotinamide in urine was not as good a measure of the low (6.1mg NE/d intake) as the individual metabolite excretions and was not sensitive to the intake of 10.1mg NE/d (see Table 20c.1). Hence, the ratio appears insensitive to marginal intakes of niacin. Moreover, some investigators suggest that excretion of both N'-methylnicotinamide and 2-pyridone in the urine is strongly dependent on the level of protein intake, so that the ratio may be a measure of protein adequacy and not niacin status (Shibata and Matuso, 1989a; Shibata and Matuso, 1989b). In view of these uncertainties, the use of this ratio is not recommended in marginal niacin deficiency states. Nevertheless, WHO (2000) include the ratio of 2-pyridone to N'-methylnicotinamide as one of the biomarkers for assessing niacin status, although they recognize that its interpretation is rather equivocal. Measurement of urinary N'-methylnicotinamide is not appropriate for pregnant women, in whom elevated excretion levels of N'-methylnicotinamide occur as a result of alterations in pyridoxine metabolism (Ftukijwatari et al., 2004).

Some of the early studies used a nicotinamide load test involving the intramuscular administration of a 50mg test dose of nicotinamide, followed by measurement of N'-methylnicotinamide in urine collected at the end of a 4-5h post-dose period. Gontzea et al. (1976) reported a 14% recovery of the test dose in well-nourished subjects over a 3h period, compared with 8% for a rural population whose basal excretion of N'-methylnicotinamide was at the lower end of the normal range. With a lower oral niacin load (nicotinamide at 20mg/70kg body weight), post-dose urinary changes in 2-pyridone were more responsive to niacin status than those changes apparent in N'-methylnicotinamide. Loading tests are impractical, however, for field surveys, so they have had very little use.

Increasingly, with the development of new analytical methods, relationships between the other urinary pyridone turnover products including N'-methyl-4-pyridone-3-carboxamide (4-pyridone) and N'-methyl-6-pyridone-3-carboxamide (6-pyridone) and niacin intakes are being examined. During an outbreak of pellagra in Mozambican refugees, for example, French researchers measured 4-pyridone and 6-pyridone as well as N'-methylnicotinamide in 24h urine collections (Dillon et al., 1992). They noted that the ratio of 6-pyridone to N'-methylnicotinamide in 24h urine specimens correlated well with the development of clinical symptoms of pellagra, mainly dermatitis, as shown in Table 20c.2.

Table 20c.2 Urinary excretion of niacin metabolites by Mozambican women. Data from Dillon et al., 1992.

Metabolite	Control (n=9)	No signs of pellegra (n=9)	
N'methyl-6-pyridone-3-carboxamide (mmol/24hr urine)	557 ± 199	152 ± 104	
N'methylnicotinamide (mmol/24hr urine)	273 ± 50	162 ± 17	
Molecular ratio of N'methyl-6-pyridone-3-carboxamide to N'methylnicotinamide	1.95	0.90	

Interpretive criteria

Interpretive criteria for adults and pregnant women for the urinary excretion of niacin metabolites expressed relative to creatinine for N'-methylnicotinamide, and as the ratio (2-pyridone / N'-methylnicotinamide) are shown in Table 20c.3. These criteria are based on early studies and have been compiled by WHO for two of their publications: Pellagra and its prevention and control in major emergencies (WHO, 2000a) and Management of nutrition in major emergencies (WHO, 2000b). As noted earlier, WHO do indicate that the interpretation of these biomarkers is equivocal.

Table 20c.3 Guidelines for the interpretation of urinary excretion of N'methylnicotinamide as mg/g creatinine in a 24hr sample and WHO, 2000b.

	Deficient	Low
N'methylnicotinamide Men; women, non-preg- nant or 1st trimester 2nd trimester 3rd trimester (mg per g creatinine in 24h urine sample)	<0.5 <0.6 <0.8	0.5-1.59 0.6-1.99 0.8-2.49
Molecular ratio of N'methyl-6-pyri- done-3-carboxamide (2-pyridone) to N'methylnicotinamide	<0.5	<1.0

Note Creeke et al. (2007) developed interpretive criteria to identify clinical pellagra based on two urinary metabolites expressed relative to creatinine from spot urine samples collected on both healthy and pellagra patients. These interpretive criteria were reported to achieve a sensitivity of 91% and a specificity of 72% for identifying clinical pellagra in their study, and are shown below:

2-pyridone: <3.0 µmol/mmol creatinine (<4.0mg/g creatinine)

N'-methylnicotinamide: >1.3 µmol/mmol creatinine (>1.6mg/g creatinine).

WHO (2000a) also provide provisional criteria for assessing the severity of public health problem of niacin deficiency based on two urinary metabolites of niacin (Table 20c.4). Note they include the ratio of 2-pyridone to N'-methylnicotinamide in Table 20c.4, even though this ratio may not be useful in populations with marginal niacin status. (Table 20c.4).

Table 20c.4 Provisional criteria for the severity of the public health problem niacin deficiency. Data from WHO, 2000b.

Indicator	Mild	Moderate	Severe
	≥ 1 clin. case, <1% of pop. in age group concerned	1-4% of pop. in age group concerned	≥ 5% of pop. in age group concerned
Urinary N'-methyl nicotinamide <0.50 mg/g creatinine	5-19%	20-49%	≥50%
Ratio 2-pyridone: N'-methylnicotin- amide <1.0	5-19%	20-49%	≥50%

Indicator	Mild	Moderate	Severe
Dietary intake of niacin equivalents <5mg/day	5-19%	20-49%	≥50%

20c.2.2 Measurement of urinary excretion of N'-methylnicotinamide and 2-pyridone

With the development of improved HPLC techniques, several urinary pyridones, including 2-pyridone, 4-pyridone, and 6-pyridone together with N'-methylnicotinamide can be readily separated and measured, with improved accuracy and sensitivity. Separation of the metabolites in urine can be achieved using reversed-phase HPLC, hydrophilic liquid interaction chromatography, normal-phase HPLC, and supercritical fluid chromatography. Detection is carried out by UV absorption, MS, or MS/MS.

Creeke & Seal (2005) have developed an HPLC method in which both N'-methylnicotinamide and 2-pyridone in urine can be analyzed in the same run. They use a polymer-based mixed mode anion exchange reverse-phase cartridge that is commercially available. Analysis is performed on a reverse-phase C18 column using a methanol gradient elution system.

20c.2.3 Niacin and niacin metabolites in plasma and erythrocytes

The concentrations of niacin compounds and niacin metabolites in the plasma, erythrocytes, and leukocytes have been studied as potential measures of niacin nutriture. Results have been inconsistent. Niacin circulates in the plasma as nicotinamide and nicotinic acid, although nicotinamide is the major form. These levels in plasma are low and reflect dietary intake rather than body stores: they do not appear to be very useful measures of niacin status (Jacob et al., 1989). Likewise, concentrations of the niacin metabolites, N'-methylnicotinamide and N'-methyl-2-pyridone-5-carboxamide (2-pyridone) in plasma do not appear to be as reliable as the same metabolites in urine for assessing individuals at risk to low niacin intakes (Jacob et al., 1989). With the development of new HPLC measurement techniques with improved accuracy and sensitivity, more research is needed to confirm these earlier findings.

The co-enzymes NAD and NADP are the active anabolic products of niacin and thus theoretically should provide a more direct measure of functional niacin status. However, in a study in Angola where niacin deficiency is endemic (Creeke et al., 2007), neither NAD and NADP concentrations in whole blood nor the NAD:NADP ratio were significantly depressed in patients with clinical pellagra (n=34). It is possible that such unexpected results were associated with inadequacies in the processing and storage of the whole blood under field conditions. This may be a factor limiting the usefulness of the whole blood NAD and NADP assay in the field.

In contrast, experimental niacin depletion - repletion studies have shown that erythrocyte concentrations of NAD (but not NADP) are sensitive to short-term changes in niacin intake, even in the absence of clinical signs of deficiency. In the depletion-repletion study in young men (Figure 20c.3),

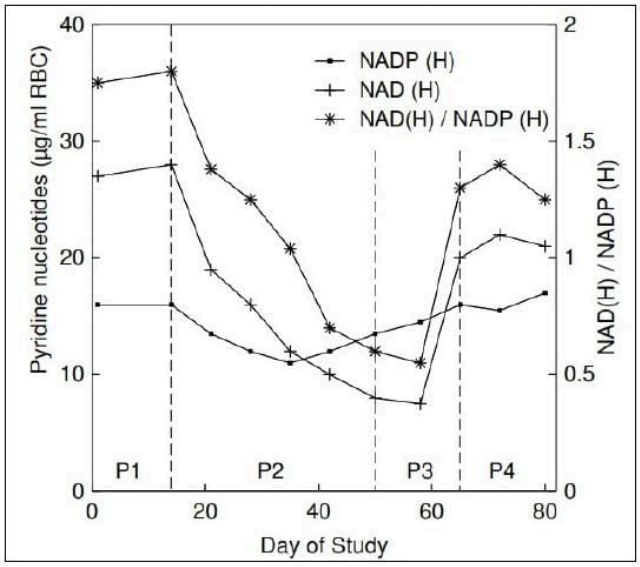


Figure 20c.3. Mean concentration of erythrocyte nicotinamide (NAD) and nicotinamde adenine dinucleotide phosphate (NADP), and the NAD/NADP ratio during four dietary periods: P1, stabilization; P2, depletion (6.1 or 10.1mg NE/d); P3, repletion; and P4, high intake (25 or 32mg NE/d). Data from Fu et al.,1989.

there was a continuous decrease in erythrocyte NAD concentrations during the depletion period, levels falling by 70% in men fed low-niacin diets containing either 6 or 10mg NE/d, increasing during the repletion phase when intakes of niacin (as NE) were adequate. In contrast, erythrocyte NADP concentrations remained unchanged. However, after five weeks of intakes of 25 and 32mg NE/d, no further significant increase in erythrocyte NAD concentrations was observed compared with the concentration during the repletion intake of 19.2mg NE/day (Fu et al., 1989). Erythrocyte NAD concentrations followed a similar trend in an experimental niacin depletion study of elderly individuals. These trends in NAD relative to NADP concentrations parallel those in fibroblasts grown in niacin-restricted cultures (Jacobson EL, 1993). Based on these results, erythrocyte NAD levels may serve as a sensitive and reliable biomarker for risk of niacin deficiency.

The ratio of erythrocyte NAD to NADP (called niacin index) is sometimes used as a measure of niacin status, as NADP content in blood remains relatively unchanged even when low-niacin intakes cause a fall in NAD concentrations, as shown in Figure 20c.3. The ratio of erythrocyte NAD to NADP of <1.0 has been set as an interpretive criterion to identify individuals at risk of niacin deficiency (Fu et al., 1989). Clearly, in populations at risk to niacin deficiency, the use of NAD and NADP concentrations in erythrocytes as well as the determination of the niacin index in whole blood should be explored further(Jacobson & Jacobson, 1997).

20c.2.4 Measurement of niacin coenzymes in erythrocytes

Accurate measurement of NAD+ in erythrocytes is technically challenging, with differences in sample processing and measurement strategies profoundly altering the results. Earlier methods were based on multiple liquid chromatography tandem mass spectrometry (MS) but the sample preparation was complicated, the run times were long, and separation of the metabolites was not always appropriate (Bustamante et al., 2017).

Demarest and co-workers (2019) have developed a rapid LC-MS/MS method for the determination of NAD+ and its metabolites and have employed this method for erythrocytes. The method has a simple, optimized sample handling procedure and can be used for erythrocytes as well as for skeletal muscle and cerebrospinal fluid. The method utilizes an Accucore HILIC column with a linear gradient with a total run time of 14min, less than half the run time of most other available methods. In addition, the sample preparation does not require drying steps or speed va, increasing the stability of NAD+.

^{1}H NMR Spectroscopy can also be used to detect and quantify NAD+ and NADP in erythrocytes (and platelets) and is claimed to be superior to mass spectrometry methods (Shabalin et al., 2018).

This page titled 20.2: Indices of niacin status (20c.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

21: FOLATE (CHAPTER 22A)

Abstract

The function of folate, in its various co-factor forms, is in mediating one-carbon metabolism, a network of pathways involving the transfer and utilization of single-carbon units, including methylene, forminino, methyl, methenyl, and formyl groups. Folate is thus essential for key biological functions, including the biosynthesis of DNA, serine and glycine metabolism, and methionine synthesis. Folate, in the form of 5-methyltetrahydrofolate, along with vitamin B_{12} (as methylcobalamin), is required for the synthesis of methionine from homocysteine, and in turn the generation of S-adenosylmethionine, a methyl group donor used in numerous reactions, including the methylation of DNA, RNA, proteins, and phospholipids. Clinical deficiency of folate is manifested as megaloblastic anemia, characterized by abnormal cell replication in the hematopoietic system, megaloblasts in the bone marrow and macrocytes in the peripheral blood. The megaloblastic anemia of folate deficiency is identical to that of vitamin B_{12} deficiency, and specific biomarker testing is essential to provide a differential diagnosis. Folate-related anemia occurs commonly in pregnant and lactating women in low- and middle-income countries. Clinical folate deficiency is less common in high-income countries, but subclinical deficiency is widespread, especially in women of reproductive age and in the presence of certain diseases and drugs. Notably, maternal folate nutrition before and in early pregnancy plays a critical role in fetal development, with conclusive scientific evidence that folic acid supplementation in early pregnancy protects against the occurrence of neural tube defects. Serum and red blood cell (RBC) folate are the biomarkers used to assess folate status, whilst plasma homocysteine provides a functional indictor of status.

21.1: Folate (22a.1)

21.2: Biomarkers of folate status (Chapter 22a.2)

21.3: Conclusions (22a.3)

This page titled 21: Folate (Chapter 22a) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

21.1: Folate (22a.1)

Folate is a generic term referring to both natural folates and folic acid (pteroylmonoglutamic acid, PGA), the synthetic form used in supplements and fortified food. All folate forms comprise three moieties: a pteridine; a p-aminobenzoic acid (PABA) and a glutamate residue (Figure 22a.1).

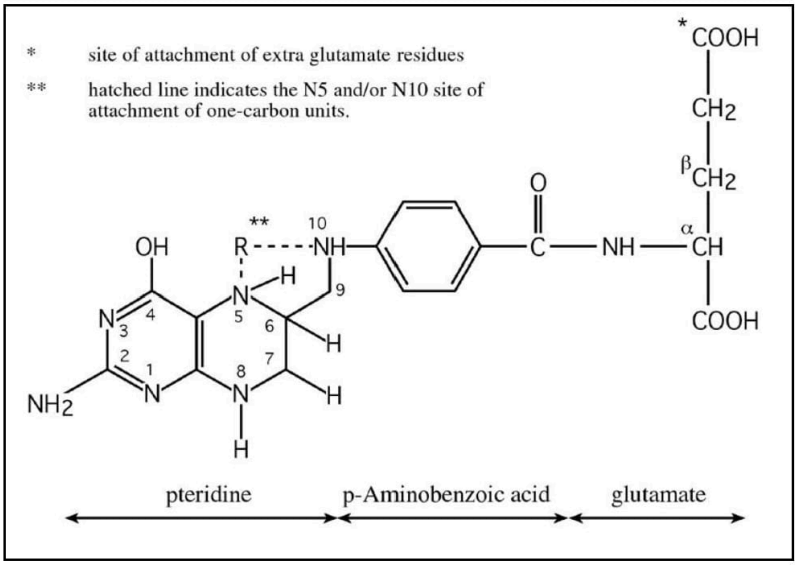


Figure 22a.1 The structure of tetrahydrofolate (THF)

The parent compound, PGA, is completely oxidized and not found in nature. The natural folate forms are reduced molecules, with the addition of 2 or 4 hydrogen atoms to the pteridine, giving rise to dihydrofolate or the various tetrahydrofolate (THF) forms. THF can carry one-carbon groups attached at the N-5 (methyl, formyl or formimino), the N-10 (formyl) or bridging N-5 and N-10 (methylene or methenyl) positions of the pteridine ring, giving rise to a number of different cofactor forms of folate. Also, whereas folic acid is a monoglutamate, containing only one glutamic acid residue, most natural food folates exist as poly-glutamate derivatives containing additional glutamate residues bound in peptide linkage to the gamma-carboxyl group.

22a.1.1 Functions of folate

Folate coenzymes are required for one-carbon metabolism involving the transfer and ulilization of single-carbon atom units, including methylene (CH_2), forminino (CH=NH), methyl (CH_3), methenyl (CH), and formyl (CHO) groups (Figure 22a.2).

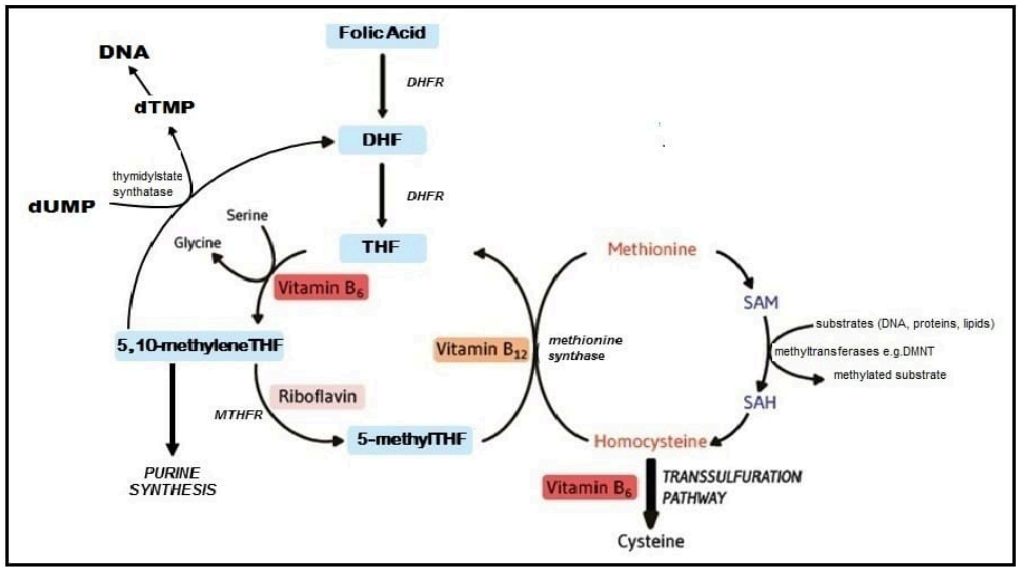


Figure 22a.2 Overview of Folate and related B vitamins in One-carbon Metabolism. Abbreviations. DHF, dihydrofolate; DHFR, dihydrofolate reductase; DMNT, DNA methyltransferase; dTMP, deoxythymidine monophosphate; dUMP, deoxyuridine monophosphate; MTHFR, methylenetetrahydrofolate reductase; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine; THF, tetrahydrofolate. Adapted from "Folate: Micronutrient Information Center, Oregon State University" (2021)

Folate is thus essential for key biological functions, including, serine and glycine metabolism, histidine catabolism, methionine synthesis and in thymidylate and purine biosynthesis, precursors of DNA. Folate, in the form of 5-methylTHF, along with vitamin B_{12}, is required for the synthesis of methionine from homocysteine (catalyzed by the enzyme methionine synthase), and in turn, the synthesis of S-adenosylmethionine (SAM). This methyl group donor is used in numerous biological methylation reactions, including the methylation of a number of sites within DNA, RNA, proteins, and phospholipids. The reader is referred to Bailey et al. (2015) for a detailed review of the functions of folate.

Folate interaction with other B vitamins. For folate to function within one-carbon metabolism, it interacts closely with vitamin B_{12}, vitamin $B_6$and riboflavin (McNulty et al., 2019). Reduced folates enter the one-carbon cycle as THF which acquires a carbon unit from serine in a vitamin B_6-dependent reaction to form 5, 10 methyleneTHF. This co-factor form, once formed either converted to 5-methylTHF or serves as the one-carbon donor in the synthesis of nucleic acids, where it is required by thymidylate synthetase in the conversion of deoxyuridine (dUMP) to deoxythymidine monophosphate (dTMP) for pyrimidine biosynthesis, or is converted to other folate forms required for purine biosynthesis. Methylenetetrahydrofolate reductase (MTHFR) is a riboflavin-dependent enzyme that catalyzes the reduction of 5, 10 methyleneTHF to 5 methylTHF, the folate form used by methionine synthase for the vitamin B_{12}-dependent conversion of homocysteine to methionine and the formation of THF. Methionine is activated by ATP to form S-adenosylmethionine (SAM), which then donates its methyl group to more than 100 methyltransferases for a wide range of substrates such as DNA, hormones, proteins, neurotransmitters and membrane phospholipids, all of which are regulators of important physiological processes (Bailey et al., 2015). In summary, effective folate functioning requires essential metabolic interaction with vitamins B_{12}, B_6 and riboflavin. Thus sub-optimal status of one or more of these B vitamins, or polymorphisms in folate genes, can impair one-carbon metabolism, even if folate status is sufficient.

Absorption and transport of folate. Folate absorption takes place by an active process, primarily from the proximal part of the jejunum. Before absorption, the polyglutamate forms of folate are deconjugated to the monoglutamate form in the gut by the intestinal brush-border enzyme glutamate carboxypeptidase II (GCPII), more commonly termed “folate conjugase”. The monoglutamates are then taken up by the mucosal cells. Folates require transporters to cross cell membranes; these include the reduced folate carrier (RFC), the proton-coupled folate transporter (PCFT), and the folate receptor proteins, FRα and FRβ. Folic acid is a monoglutamate and thus does not require deconjugation before uptake by intestinal cells.

Most of the folate monoglutamates absorbed from the gut are transported to the liver (the major storage organ for folate) and are reconverted intracellularly within hepatocytes to polyglutamate derivatives by the enzyme folylpolyglutamate synthetase.

These folate polyglutamates are stored in the liver or converted to 5-methyltetrahydrofolate for secretion into the bile, and then reabsorbed by way of the enterohepatic circulation. This recirculation process may account for as much as 50% of the total folate that reaches the peripheral tissues. Circulating folate in blood is found in the monoglutamate form, predominantly as 5-methyltetrahydrofolate.

22a.1.2 Deficiency of folate in humans

Severe deficiency of folate leads to megaloblastic anemia, presenting as fatigue, weakness, and shortness of breath owing to a low red blood cell count. This condition is hematologically characterized by the presence of immature, enlarged nucleated cells, reflecting impaired DNA synthesis as a result of folate depletion. Rapidly proliferating cells are especially sensitive to abnormalities in DNA synthesis. Hence, the manifestations of folate deficiency appear first in the hematopoietic system and then in the epithelial cell surfaces and the gonads. Abnormal cell replication in the hematopoietic system, manifested by hypersegmented neutrophils, is one of the earliest morphological changes. Later, megaloblasts appear in the bone marrow and macrocytes in the peripheral blood. Some additional signs and symptoms that have been reported with clinical folate deficiency include fatigue, angular cheilosis, anorexia, insomnia, glossitis, recurrent aphthous ulcers, and pallor of the skin and mucous membranes. Of note, the megaloblastic anemia resulting from folate deficiency is identical to that resulting from vitamin B_{12} deficiency, and therefore specific biomarker testing is essential to provide a differential diagnosis (See Section 22a.2).

Causes of low and deficient folate status. Folate deficiency arises from various causes, relating either to increased requirements, reduced availability, or both. Pregnancy is a time when folate requirement is greatly increased in order to sustain the demand for rapid cell replication and growth of fetal, placental and maternal tissue (McNulty et al., 2019). Gastrointestinal conditions, such as celiac disease, can also lead to deficient folate status through chronic malabsorption. Certain drugs, including phenytoin and primidone (anticonvulsants) and sulfasalazine (used in inflammatory bowel disease), are also associated with folate deficiency through adversely affecting folate metabolism. Likewise, heavy alcohol consumption and smoking are also linked with lower folate status (Bailey et al., 2015).

Prevalence of deficiency. Folate deficiency occurs commonly in pregnant and lactating women in low- and middle-income countries, where it is reported to occur in > 20% of women of reproductive age (Rogers et al., 2018). Here, dietary intakes of folate are often inadequate to meet the high requirements of pregnancy.

In high-income countries, clinical folate deficiency is less common (due largely to the beneficial effects on folate status of supplementation and/or food fortification), but subclinical deficiency (indicated by low serum and RBC folate concentrations) is widespread, particularly in women of reproductive age, especially if pregnant and lactating, in European countries. Low folate status has also been reported in low birth weight and premature infants (Scholl & Johnson, 2000) and adolescents of low socioeconomic status (Bailey et al., 2015). Serum and RBC folate concentrations typically decrease throughout pregnancy; however, supplementation with folic acid prevents this decline and can thus prevent the occurrence of megaloblastic anemia of pregnancy (Blot et al., 1981; McNulty et al., 2013). Estimates of folate deficiency can however vary considerably across different populations and population sub-groups, depending on the method used and the cutoff points applied to biomarker measures of folate status (Bailey et al., 2015).

Neural tube defects. Maternal folate nutrition before and in early pregnancy is known to play a critical role in fetal development. Notably, conclusive scientific evidence published 30y ago shows that folic acid supplementation with folic acid (the synthetic form of the vitamin) in early pregnancy protects against both first occurrence (Czeizel & Dudás, 1992) and recurrence (MRC, 1991) of neural tube defects (NTDs). These major birth defects occur as a result of a failure of the neural tube to close properly in the first few weeks of pregnancy, leading to death of the fetus or newborn, or to various disabilities involving the spinal cord, the most common form of which is spina bifida. The conclusive evidence that folic acid can prevent NTD has led to clear folic acid recommendations for women of reproductive age which are in place worldwide. It is important to appreciate, however, that the risk of NTD is increased when maternal folate status is low, but not necessarily within the range typically classed as folate deficiency. In a large prospective study of women in Ireland (where rates of NTD are among the highest in the world), a woman's risk of having a child with an NTD was found to be strongly associated with pregnancy RBC folate in a continuous dose-response relationship (Daly et al., 1995).

In practice, implementing folic acid recommendations so that women and their babies benefit, is somewhat challenging. Although mandatory folic acid fortification of foods has proved to be highly effective in reducing NTD wherever it has been introduced (in 84 countries worldwide), elsewhere preventable NTDs are not being prevented including in European countries. Notably, one recent study estimated that from 1998 to 2017, a total of 95,213 NTD pregnancies have occurred amongst 104 million births in

28 European countries; a prevalence of 0.92 per 1,000 births (Morris et al., 2021). This study concluded that failure to implement mandatory folic acid fortification in the 28 European countries continues to cause NTD to occur in almost 1,000 pregnancies every year.

The precise mechanism explaining the beneficial effects of periconceptional folic acid against NTD remains uncertain, though proposed mechanisms have focussed on factors that impair normal folate metabolism, including polymorphisms in folate genes. Among the latter, an increased risk of NTD is most strongly associated with the 677C>T variant in the gene encoding the folate-metabolising enzyme methylenetetrahydrofolate reductase (MTHFR), as reported in most studies and meta-analyses (Botto & Yang, 2000; Vollset & Botto, 2004). Autoantibodies against folate receptors are also implicated in pregnancies affected by NTD (Rothenberg et al., 2004). Also, although low maternal folate is considered the major contributing factor in NTD, convincing evidence shows that low vitamin B_{12} is an independent risk factor in NTD (Molloy et al., 2009). In addition, apart from preventing NTD, there is good evidence that periconceptional folic acid use may prevent congenital heart defects in infants (van Beynum et al., 2010), and possibly orofacial clefts (Bailey et al., 2015).

22a.1.3 Food folate sources, bioavailability and dietary intakes

Folates are widely distributed in animal and plant foods, but only certain foods are rich sources, including liver, yeast, green leafy vegetables, asparagus, beans, legumes, where they exist primarily as polyglutamates, containing up to nine glutamate residues attached to the *p*-aminobenzoic group of the molecule (Figure 22a.1). In contrast, folic acid, the synthetic vitamin, is a mono-glutamate, containing only one glutamic acid residue in its structure. Also, unlike folic acid, which is a fully oxidized molecule, natural folates are reduced at the 5, 6, 7 and 8 positions of the pteridine ring and so are prone to oxidative cleavage at the C9-N10 bond producing two degradation products, a pteridine and *p*-aminobenzoylglutamate, both of which are inactive and cannot be biologically converted to any active folate form.

Folate bioavailability. Bioavailability refers to the proportion of ingested nutrient that is absorbed and available for metabolic processes. Naturally-occurring food folates are only partially bioavailable (McNulty & Pentieva, 2004). Natural food folates are reduced molecules, and therefore are inherently unstable outside living cells. In addition, the ease with which folates are released from different food matrices and their conversion to the monoglutamate form before uptake by intestinal cells can vary greatly. Folate bioavailability from different food sources is also dependent on the presence of certain dietary constituents, that may enhance folate stability during digestion, or inhibit bioavailability owing to specific inhibitors of deconjugation. Thus, the bio-availability of food folates from a mixed diet is limited and highly variable. In addition, natural food folates (particularly green vegetables) can be unstable during cooking, and this can substantially reduce the folate content of a food before it is ingested (McKillop et al., 2002). This is an additional factor that further contributes to the limited potential for natural food folates to positively influence folate status (Cuskelly et al., 1996).

Dietary folates. In contrast to the natural folate forms which have limited stability and bioavailability, folic acid provides a highly stable and bioavailable vitamin form. The bioavailability of folic acid is assumed to be 100% when ingested as a supplement, while folic acid in fortified food is estimated to have about 85% the bioavailability of supplemental folic acid (Pfeiffer et al., 1997). These differences have led to the development of "dietary folate equivalents" or DFE values; more details are given in Section 8a. Briefly, expressing dietary folate intakes and recommendations in DFE terms, allows an adjustment for the differences in bioavail-ability between natural food folates and the synthetic vitamin. The DFE is defined as the quantity of natural food folate plus 1.7 times the quantity of folic acid in the diet; this definition is based on the assumption that the bioavailability of folic acid added to food is greater than that of natural food folate by a factor of 1.7 (Institute of Medicine, 1998). This estimation is largely dependent on a metabolic study in non-pregnant women that estimated the bioavailability of food folates to be 50% relative to that of folic acid (Sauberlich et al., 1987), and other evidence, mentioned above, showing that folic acid added to food has 85% of the bio-availability of free folic acid (Pfeiffer et al., 1997).

Fortified foods as a source of folate. Food fortification, the process of adding essential micronutrients to food, plays an important role in facilitating a more optimal nutritional status in individuals and populations. In the case of folate, it plays a crucial role. Folic acid, the folate form used for food fortification, is cheap to produce, very stable once added to foods and highly bioavailable when ingested. Thus, depending on national fortification policy and/or access to folic acid-fortified foods, the folate status of populations can vary greatly from one country to the next This is reflected in differences in health outcomes (notably in relation to NTD risk). Food fortification may be undertaken on a voluntary or mandatory basis. Voluntary fortification, whereby folic acid is added to foods such as breakfast cereals at the discretion of the manufacturer, is permitted in most European countries. This results in higher folate intake and status (Hopkins et al., 2015), but the benefit is limited only to consumers who choose to eat the fortified food products. However, when fortification is undertaken on a mandatory (population-wide) basis, including in

the USA, Canada, Australia and Chile, it has proved to be highly effective, not only in increasing folate status and reducing folate deficiency (Yang et al., 2010), but also in reducing NTD in that country (Honein et al., 2001, Lopez-Camelo et al., 2005, De Wals et al., 2007, Sayed et al., 2008). Mandatory fortification is now in place in 85 countries worldwide, both high- and low-middle income countries

Effects of high intakes of folate. High folate intakes are generally not associated with any adverse effects. However, there are concerns of potential adverse effects of excess intakes of folate acid, the synthetic vitamin form. Excessive folic acid intake constitutes exposure doses that exceed the Tolerable Upper Intake Level (UL) of 1000µg/d for adults, as set by the U.S. Institute of Medicine (1998).

Historically the concern regarding excess folic acid focused on the potential to mask pernicious anemia and exacerbation of the clinical effects of vitamin B_{12} deficiency. More recently other concerns have been raised, including potential adverse effects on cancer risk, birth outcomes, and other diseases. However, a recent report from a 2019 expert workshop tasked with reviewing this research area, concluded that there is an insufficient body of evidence to support adverse human health outcomes as a result of high intakes of folic acid. Nonetheless, these experts called for further research to determine the safety of excess folic acid intake (Maruvada et al., 2020).

This page titled 21.1: Folate (22a.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

21.2: Biomarkers of folate status (Chapter 22a.2)

In 2015, an expert international panel produced a comprehensive review of folate biology and biomarkers as part of the Biomarkers of Nutrition for Development (BOND) project convened by the U.S. National Institutes of Health (Bailey et al., 2015). This review identified serum folate, red blood cell (RBC) folate and plasma homocysteine as the "Priority Folate Biomarkers" for assessing folate status. Serum folate reflects recent dietary intake. RBC folate, compared with serum folate, is a better indicator of folate intake and status over the previous 3–4mos. Plasma homocysteine provides a functional indicator of folate status and is elevated in folate deficiency, on the basis that normal homocysteine metabolism requires an adequate supply of folate. Concentrations of folate in erythrocytes, but not serum, also fall in vitamin B_{12} deficiency. Ideally, both serum and RBC folate concentrations should be measured.

Cut-off values for assessment of folate status. The sequential stages in the development of folate deficiency were originally established through detailed depletion / repletion experiments conducted during the late 1960's, providing a basis for setting cut-off values for serum and RBC folate for the assessment of status (Herbert, 1987); these sequential stages of folate deficiency and related cut-off values are summarized in Box 22a.1

Box 22a.1. Cut-off values for assessment of folate status in populations

Historical perspective

During the late 1960's, cut-off values for sequential stages of folate deficiency were established through depletion/repletion experiments (Herbert, 1987):

- A serum folate value < 7nmol/L (3ng/mL) indicated negative folate balance
- For RBC folate, cut-off values were set at:
  - < 363nmol/L (160ng/mL) to indicate the onset of folate depletion,
  - < 272nmol/L (120ng/mL) to mark the beginning of folate-deficient erythropoiesis, and
  - < 227nmol/L (100ng/mL) to indicate folate-deficient anemia

More recently
Folate deficiency was defined more simply, using a single cut-off value for each biomarker, based on a metabolic indicator (increased plasma homocysteine) (Selhub et al., 2008):

- Serum folate < 10nmol/L to indicate folate deficiency
- RBC folate < 340nmol/L to indicate folate deficiency

These cutoff values have been recommended by the 2005 WHO Technical Consultation on folate and vitamin&bsp; B_{12}deficiencies for the assessment of folate status of populations (de Benoist, 2008). Adapted from: Bailey et al., 2015;Herbert, 1987; Selhub et al., 2008; de Benoist, 2008.

The initial stage of folate deficiency, termed negative folate balance, is consistent with serum folate values of < 7nmol/L (Herbert, 1987). If the negative balance persists, tissue folate becomes depleted, as indicated by RBC folate falling below the normal range to < 363nmol/L. At this second stage, there is little evidence that biochemical function is impaired, although plasma homocysteine concentrations may be slightly elevated. By the third stage, termed folate-deficient erythropoiesis, functional impairment is usually evident, with erythrocyte folate values < 272nmol/L. Tissue folate stores are severely depleted in the fourth and final stage where the classical hematological changes occur, manifested as folate-deficiency anemia. These include macro-ovalocytic erythrocytes in the circulating blood and megaloblasts in the bone marrow. At this stage, hypersegmented neutrophils in the peripheral blood smear and abnormal RBC indices are also apparent; mean red cell volume and mean cell hemoglobin are elevated and hemoglobin concentration is low.
More recently, revised cut-off points for folate deficiency (serum folate <10nmol/L and RBC folate < 340nmol/L) have been defined based on the metabolic indicator, plasma homocysteine (Selhub et al., 2008), as summarized in Box 22a.1.

22a.2.1 Serum and RBC folate concentrations

Serum folate represents the sum of several folate forms circulating in blood. The main circulating folate is 5-methyl-THF; other reduced forms such as THF and formyl-THF may also be present, but in very small concentrations (Bailey et al., 2015). As will be discussed later (section 22a.2.4), unmetabolized folic acid can also be present in varying concentrations in plasma (Bailey et al., 2010).

Red cells contain much higher folate concentrations than serum. 5-methyl-THF polyglutamates are the main RBC folate forms. The measurement of RBC folate is more complex than that of serum folate, because of the need to convert polyglutamates to mono-glutamates prior to analysis. In individuals homozygous for the *MTHFR* C677>T polymorphism, a portion of the 5-methyl-THF polyglutamates is replaced by formyl-folates (Bagley & Selhub, 1998).

Serum folate is the earliest indicator of altered folate exposure and reflects recent dietary intake (Pfeiffer et al 2010). In the individual, serum folate concentrations can increase markedly in response to dietary folate, reaching a peak concentration 90min after ingestion.

RBC folate is a sensitive indicator of long-term folate status. RBC folate parallels liver concentrations, accounting for about 50% of total body folate, and thus reflects tissue folate stores (Wu et al., 1975). RBC folate, compared with serum folate, responds more slowly to changes in dietary folate intake and is a better indicator of folate intake over the previous 3–4mos when circulating folate is incorporated into the maturating red cells during erythropoiesis, thereby reflecting folate status during the preceding 120d, the half-life of the RBC (Mason, 2003).

Factors affecting serum and RBC folate concentrations

Dietary folate intakes. Blood folate levels are primarily affected by dietary folate intakes. Serum and RBC folate are highly responsive to intervention with folic acid (Duffy et al., 2014), with natural food folates typically resulting in poorer folate responses compared to folic acid at similar intervention levels (Cuskelly et al., 1996). Likewise, population data show that both serum and RBC folate concentrations are highly reflective of exposure to folic acid, with the highest concentrations observed in people who consume folic acid in both supplements and fortified foods (Yang et al., 2010; Hopkins et al., 2015).

Vitamin B_{12} status. Because vitamin B_{12} is required for normal folate recycling and folate retention within cells, vitamin B_{12} deficiency leads to a failure to retain folate within cells (Hoffbrand & Weir, 2001). Consequently, RBC folate concentrations fall, despite the presence of normal (or sometimes even elevated) serum folate. Thus, low RBC folate may reflect vitamin B_{12} deficiency as well as folate deficiency.

Fasting versus non fasting samples. Samples from fasted individuals, on average, have lower concentrations of serum (by 10%) and RBC folate (by 5%) compared with samples from nonfasted (< 3h) participants, but the small differences generally indicate that fasting is not essential when assessing the folate status of populations. (Bailey et al., 2015).

Important preanalytical factors. Detailed information on preanalytical factors affecting serum and RBC folate is described in a comprehensive review article on analytical approaches by Pfeiffer et al., 2010. Folate is the least stable of the B vitamins and is susceptible to oxidative degradation during preanalytical sample handling and storage. Careful sample handling and use of antioxidants are therefore required to maintain sample integrity. Blood for serum folate analysis should be processed and frozen promptly. Although serum is generally preferred over plasma in most laboratories, both matrices generally produce comparable results for folate, as long as the sample processing is not delayed. If delays are unavoidable, the sample should be protected from light, kept cool and processed within a few days of collection. Folate in serum and hemolysates (but not in whole blood) can withstand a few short freeze / thawing cycles, particularly if the vials are kept closed as much as possible to minimize the exposure of the sample to air. Folate in serum / plasma degrades rapidly at room temperature, particularly in the presence of EDTA.

Analytical methodologies for measurement of serum and RBC folate

Over the past 50y, analytical methods to assess serum and RBC folate concentrations have been continuously improved; however, they have not yet reached the point where they produce sufficiently comparable results across methods or laboratories. The within-person variability for serum folate is about twice that for RBC folate (CV of 21.5% and 9.1%, respectively, Bailey et al., 2015).

Critical considerations in folate assessment. Assessing folate status is complicated by the large number of folate forms that may be readily interconverted. To overcome this problem, microbiological assays have been used for many decades due to the ability of some bacteria to grow in the presence of many different forms of folate, i.e. *L.rhamnosus* responds to all active monoglutamate forms (see below). Subsequently, assays using competitive protein binding became common because of their simplicity. In more recent years, the use of LC-MS/MS has emerged. Of note, the emergence of LC-MS/MS enables investigation of relevant research findings related to folate metabolism. These include alterations in the relative proportion of different folate forms in red cells owing to the common *MTHFR* C677>T polymorphism, along with an overall lower RBC folate (Molloy et al., 1997; Bagley & Selhub, 1998; Shane et al., 2011), and the presence of free ("unmetabolized") folic acid in the blood as a result of high intakes of folic acid (Pfeiffer et al, 2010).

There are three main method types for measurement of serum and RBC folate, each with advantages and disadvantages (Bailey et al., 2015).

Microbiological assay. This is widely considered to be the gold standard assay for serum and RBC total folate because it measures all biologically active forms equally and does not measure folate species that lack vitamin activity. The underlying principle of the microbiological assay is that a folate-dependent microorganism, namely *Lactobacillus rhamnosus* (formerly called *Lactobacillus casei*), grows proportionally to the amount of folate present in serum or whole blood and the folate concentration can be quantified by measuring the turbidity of the inoculated medium after a 2d incubation. A chloramphenicol-resistant strain of *L.rhamnosus* is used and the assay is performed using automated microtiter plate technology.

The key advantages of the microbiological assay include its excellent sensitivity, low cost and simple instrumentation (thus suited for low-resource settings), and the fact that it can be used with dried blood spot samples. The high sensitivity is a particular advantage when limited volume is available, such as for samples collected from a fingerstick or dried blood spot. The disadvantages are that the microbiological assay is relatively laborious unless automated liquid handling is introduced, lengthy assay time with limited throughput and limited linear range (thus requiring dilution of samples). It is also prone to contamination issues and potential interference by the presence of antibiotics or antifolates. The latter limitation is well recognized, but in practice may not be such an issue, in that analysts in the US reported that < 1% of samples from the population-based NHANES cohort exhibited a pattern of interference due to the presence of antibiotics or antifolates (Pfeiffer et al, 2010).

Protein-binding assays. These assays were developed primarily for clinical settings, to enable the diagnosis of folate deficiency. Protein-binding assays use the highly specific folate-binding protein to extract folate from the sample. The strengths of this approach include the high sample throughput, quick turnaround time, availability in commercial kit form and minimum operator involvement. The disadvantages are that the various folate forms have different affinities to folate-binding protein, the questionable accuracy when mixtures of folate are present, the limited linear range (thus requiring dilution of samples), matrix effects when sample is diluted, and lot-to-lot variability of commercial kits. Also, although not sensitive to antibiotics, protein-binding assays are influenced by certain antifolates such as methotrexate.

Chromatography-based assays (HPLC-FD, LC-MS/MS). Chromatography-based methods typically provide information on individual folate forms based on measurement of intact folates via HPLC. More recently, LC-tandem mass spectrometry (LC-MS/MS) is now the preferred detection for HPLC-based methods in specialized laboratories. Advantages of chromatography-based approaches are that they measure all folate forms, are highly selective and specific, have good sensitivity and precision, enable in-house control of performance and use of stable-isotope-labeled internal standards to compensate for procedural losses. Ensuring accurate calibration is a big task for chromatography-based methods, due to the high number of folate forms and also because of their instability. Disadvantages include the high costs and requirement for expensive instrumentation, experienced operator and frequent technical service, along with being a relatively laborious approach unless automated liquid handling is introduced, and the requirement for complex sample extraction / cleanup. Also interconversions of folate forms during the assay procedure need to be considered in interpretation of data.

Cut-off values and Interpretation of Serum and RBC folate

The measurement of total folate provides information on the folate status of the individual, either in the short-term through serum folate, or in the long-term through RBC folate. An historical perspective on the use of folate cutoffs is provided in Box 22a.1. However, the inconsistent use of cutoff values over time to assess the proportion of populations with deficient or low folate status led to a certain degree of confusion. More recently, revised cut-off points for folate deficiency (serum folate < 10nmol/L and RBC folate < 340nmol/L) have been defined based on the metabolic indicator, plasma homocysteine (Selhub et al., 2008), as summarized in Box 22a.1. These cutoff values have been recommended by the WHO for assessing folate status of populations (de Benoist, 2008). Of note, the values were derived from data generated using the microbiological assay.

As discussed earlier, the risk of NTD is associated with a maternal folate status that would not conventionally be classed as deficient. Although there is no cutoff value established by international organizations for folate concentrations to define NTD risk in populations, it is accepted that the number of NTDs that can be prevented in a population is dependent on maternal folate status. Specifically, RBC folate concentration in the mother has been shown to be a sensitive biomarker of NTD risk. In the one and only prospective study that has been conducted to date, Daly et al. (1995) found that the prevalence of NTD in an Irish population was lowest when maternal RBC folate concentrations were ≥ 906nmol/L (400ng/mL). Subsequently, data modeled from folic acid intervention studies in China by Crider et al. (2014) confirmed the dose response relationship between RBC folate concentrations and NTD risk as shown in the Irish study (Daly et al., 1995).

22a.2.2 Plasma homocysteine concentrations

Plasma homocysteine provides a sensitive functional indicator of folate status. On the basis that normal homocysteine metabolism requires an adequate supply of folate, plasma homocysteine becomes elevated when folate status is low.

Homocysteine is a four-carbon, thiol-containing amino acid found in human plasma, mostly present in the form of various disulfides, such as homocysteine-cysteine disulfide. About 75% of total homocysteine is bound to protein (mainly albumin), whereas the remainder occurs in nonprotein-bound "free" forms. Total homocysteine is defined as the sum of all homocysteine species in serum or plasma, including free and protein-bound forms. Only a very small portion (1–2%) of plasma homocysteine is present as the thiol, however the relative contribution of the thiol to total homocysteine increases to 10–25% in patients with abnormally elevated plasma homocysteine (Bailey et al., 2015).

Homocysteine is derived from the essential amino acid, methionine. It is metabolized in two ways: remethylation to methionine (by methionine synthase) or transsulfuration to cystathionine (by cystathionine beta-synthase; CBS) then to cysteine (Figure 22a.2). These pathways require adequate status of folate and the metabolically related B-vitamins. The remethylation pathway, whereby methionine is synthesized from homocysteine by methionine synthase, is dependent on both folate and vitamin B_{12} as cofactors, whilst the transsulfuration pathway that converts homocysteine to cysteine is catalyzed by two vitamin B_6 dependent enzymes. A fourth B vitamin, riboflavin, is required in its cofactor form flavin adenine dinucleotide (FAD) for the activity of MTHFR, the enzyme that catalyzes the reduction of 5,10 methyleneTHF to 5 methylTHF. Once formed, 5 methylTHF is used by methionine synthase for the vitamin B_{12}-dependent conversion of homocysteine to methionine and the formation of THF. Thus, the concentration of homocysteine in plasma / serum is regulated by up to four B-vitamins: folate, vitamin B_{12}, vitamin B_6 and riboflavin.

Apart from providing a functional biomarker of folate, higher plasma homocysteine is associated with a number of chronic diseases of ageing, including an increased risk of cardiovascular disease (CVD; Graham et al., 1997), cognitive impairment and dementia (Smith & Refsum, 2016). It remains to be established, however, whether plasma homocysteine *per se* a risk factor for these or other diseases. In the case of CVD, this issue is particularly controversial. Despite strong and consistent evidence from observational studies over many years, several secondary prevention trials published between 2004 and 2014 failed to demonstrate a benefit of homocysteine-lowering therapy against the recurrence of CVD events in patients with existing disease. The evidence is however stronger for the relationship of homocysteine with stroke than heart disease, with good evidence from both population data and randomized trials that folic acid intervention and/or homocysteine-lowering can significantly reduce the risk of stroke, and particularly so in people with no previous history of stroke (McNulty et al., 2017). Of note, although the literature in this area focuses on homocysteine as the CVD risk factor, it is possible that folate and related B-vitamins have roles in CVD that are independent of their homocysteine-lowering effects. Thus, plasma homocysteine may be linked with CVD as a functional marker of low B-vitamin status which reliably reflects perturbed one-carbon metabolism, rather than being causatively related to CVD *per se* (McNulty et al., 2017).

Factors affecting homocysteine concentrations

Folate status. On the basis that normal homocysteine metabolism requires an adequate supply of folate, plasma homocysteine is first and foremost affected by folate status. Thus, plasma homocysteine was shown to respond within 3–4wks of folate depletion (increase) and subsequent repletion (decrease) in a controlled metabolic study in healthy women (Jacob et al., 1998). Likewise, in observational studies, plasma homocysteine is invariably found to be inversely associated with folate status, whether measured as serum or RBC folate.

Intervention with folic acid. Plasma homocysteine is highly responsive to intervention with folic acid (the synthetic vitamin form), alone or in combination with vitamin B_{12}, vitamin B_6, riboflavin and betaine (or choline). Thus, food fortification with

folic acid has marked effects on homocysteine concentrations. Using population-based data from the US, Pfeiffer et al. (2008) reported a 10% decrease in plasma homocysteine when comparing values pre-fortification (1991–1994) to post-fortification (1999–2004) of food with folic acid, as implemented on a mandatory basis in 1996–1998. Folic acid-fortification on a voluntary basis (i.e. added to food at the manufacturer's discretion) also affects plasma homocysteine. In a convenience sample of nearly 500 healthy adults in Northern Ireland aged 18–92y, who were not taking folic acid supplements, homocysteine concentrations were lower by 2µmol/L in high consumers compared to non-consumers of fortified foods (providing > 100µg/d and 0µg/d folic acid, respectively) (Hoey et al., 2007).

Folate-related B vitamins. Because normal homocysteine metabolism is dependent on four B vitamins, homocysteine concentrations will be elevated with other B vitamin deficiencies apart from folate, notably vitamin B_{12} (Allen et al., 2018). Thus whilst plasma homocysteine is primarily a folate biomarker, once folate status is optimized, a much greater dependency on vitamin B_{12} emerges (Quinlivan et al., 2002). Likewise, in population groups who consume folic acid-fortified foods or folic acid supplements, homocysteine is considered to be a more reliable biomarker of vitamin B_{12} than folate (Refsum et al., 2004).

Disease and lifestyle. Apart from inversely reflecting folate and related B vitamin status, plasma homocysteine is found to be elevated in patients with impaired renal function (Yetley et al., 2011; Allen et al., 2018) and with certain drug treatments (Refsum et al., 2004). Alcohol intake, coffee-drinking and smoking are also associated with higher plasma homocysteine.

Age, sex and lifecycle stage. Homocysteine concentrations increase throughout life and are higher in males compared to females at all ages. Although plasma homocysteine decreases in pregnancy (by about 50%), concentrations are reported to normalize within a few days postpartum (Murphy et al., 2004).

Ethnic and genetic effects. Plasma homocysteine may also differ among ethnic groups, but these differences appear to be less important than the influence of B vitamin status. Notably, the most common genetic cause of elevated homocysteine in the general population is homozygosity for the MTHFR C677>T polymorphism (affecting an estimated 10% of people worldwide but much higher in some populations, including Mexico and Northern China). This folate gene variant also contributes to a higher risk of blood pressure throughout the lifecycle, particularly when combined with deficient status of riboflavin, which is required as a cofactor for MTHFR, a key folate metabolizing enzyme (Psara et al., 2020; Ward et al., 2020). Notably, both phenotypes associated with this common polymorphism (high homocysteine and high blood pressure) appear to be modifiable with better riboflavin status (McNulty et al., 2020). Most other genetic polymorphisms in enzymes related to one-carbon metabolism have little effect on homocysteine concentrations (Refsum et al., 2004).

Inborn errors of metabolism. Homocystinuria refers to rare inborn errors of metabolism which lead to severely increased homocysteine concentrations in plasma, usually > 100µmol/L, along with large amounts excreted in urine. The most common cause of homocystinuria is deficiency of the enzyme cystathionine β-synthase (CBS). CBS deficiency, an autosomal recessive condition, has a reported worldwide birth prevalence of 1 in 344,000, while that in Ireland the frequency is much higher, at 1 in 65,000, based on newborn screening and cases detected clinically. These patients have a high risk of premature, frequently fatal, thromboembolic events. Early diagnosis and treatment with pyridoxine and/or folic acid and betaine, preferably from infancy, can however prevent CVD events and most of the clinical symptoms (Refsum et al., 2004).

Analytical methodologies for measurement of plasma homocysteine

Plasma total homocysteine is a very stable analyte as long as the plasma is separated from the red blood cells within 1hr of blood collection (or within 8 hours if the whole blood is kept on ice). As described in detail by Refsum et al. (2004), various method types are available for homocysteine determination. These range from fully automated commercial kits (immunoassay or enzymatic methods) to chromatographic assays with mass spectrometry detection, overall providing comparable results and good assay performance. All methods require the reduction of the disulfide bonds to allow measurement of total homocysteine. The reported within-person variability CV for plasma homocysteine is 12.2% (Refsum et al., 2004).

Choice of method. Because the measurement of plasma homocysteine produces comparable results across different method types, the choice of method is mainly dependent on available instrumentation and technical expertise. Bailey et al. (2015) put forward the following recommendations to help in making this decision:

- Using commercial kits (either immunoassay or enzymatic assay) on a fully-automated clinical analyzer will provide the highest throughput and quickest turnaround time with the least effort. The main disadvantage is the relatively high reagent costs, which can make the measurements quite expensive, particularly for a large number of samples.
- If an HPLC system with fluorescence detector is available, setting up a manual assay may be the least expensive approach, particularly in settings with low labor costs. The disadvantage of the manual HPLC assay is that the number of samples that

can be run is quite limited (typically to about 50–70 unknown samples per analytical run).

- If a laboratory has access to an existing LC-MS/MS system and the required technical knowhow, plasma homocysteine can be measured in a high throughput semi-automated manner, with quite low reagent costs. Due to the high initial cost, however, an LC-MS/MS system is economical only if a large number of samples are measured regularly.
- GC-MS or GC-MS/MS have also been used in high throughput research settings, particularly when the co-determination of methylmalonic acid and methionine is of interest.
- GC-based methods provide better precision, higher resolution, and longer column life compared to LC-based methods and are not subject to ion suppression issues, which can be a problem in LC-MS/MS methods.

Cut-off values and Interpretation

In one of the most comprehensive reviews of all relevant aspects of homocysteine measurement, it is proposed that reference values for homocysteine are established for different populations to account for the influence of both nonmodifiable and modifiable factors (Refsum et al., 2004; Table 22a.1).

Table 22a.1. Upper reference limits for plasma homocysteine (μmol/L) in populations. Adapted from: Refsum et al., (2004) with Data Norway, the UK and Israel and from 9 European countries participating in the European Concerted Action Project (Graham et al., 19

[1]Individuals eating folic acid-fortified food or taking folic acid-containing supplements.

Population sub-group	Folic acid Supplemented[1]	Non-supplemented
Pregnancy	> 8	> 10
Children < 15y	> 8	> 10
Adults 15–65y	> 12	> 15
Adults > 65y	> 16	> 20

The following recommendations were put forward for setting reference ranges:

- Each laboratory should establish reference limits for its region.
- The reference population should exclude individuals with folate or cobalamin deficiency or increased creatinine, and patients with diseases or who are taking drugs that increase homocysteine concentrations.
- Separate reference limits for children, adults, the elderly, and pregnant women should be used (Table 22a.1).
- In populations consuming folic acid supplements or folic acid-fortified foods, the upper reference limit is usually 20–25% lower than in nonfortified populations (Table 22a.1).
- The upper homocysteine reference limit should be interpreted with age, sex, and other determinants in mind, particularly prevailing food fortification practices.

22a.2.3 Other biomarkers of folate status

Neutrophil lobe count

Usually, neutrophils have three or four segments, but in megaloblastic anemia (owing to folate or vitamin B_{12} deficiency), this number increases. Neutrophil hypersegmentation can be evaluated in smears of peripheral blood or in white blood cells obtained from the buffy coat (i.e. interface between the serum / plasma and the sedimented red cells), and may be the earliest morphological change to appear in the blood in folate and B_{12} deficiency. Neutrophil hypersegmentation however is not specific to folate or B_{12} deficiency as it also occurs in other conditions including uremia, myeloproliferative disorders, myelofibrosis and as a congenital lesion in approximately 1% of the population, even when the status of folate and vitamin B_{12} is adequate. The observed greater incidence of neutrophil hypersegmentation in iron deficiency anemia concurrent with folate and/or vitamin B_{12} deficiency implies interactive effects on the erythropoietic process (Westerman et al., 1999; Metz, 2008).

Serum folic acid

Unmetabolized (or “free”) folic acid can be present in varying concentrations in plasma and is typically associated with higher folic acid intake via fortified foods, supplements, or a combination of both (Bailey et al., 2010). Whilst greater concentrations of

unmetabolized folic acid are generally associated with higher serum folate concentrations, there is large variation in reported values and no clear dose response relationship exists between folic acid intake and unmetabolized folic acid in plasma (Pfeiffer et al., 2004; Bailey et al., 2010; Obeid et al., 2011). Thus, cut-off values or desirable ranges have yet to be identified. One randomized trial addressed concerns relating to the potential for folic acid to exert adverse biological effects in pregnancy, and demonstrated that folic acid supplements taken at recommended levels throughout pregnancy do not lead to increased circulating unmetabolized folic acid in mothers or their babies (Pentieva et al., 2016). However, the biological impacts and the significance for health in general of unmetabolized folic acid in blood remain to be elucidated.

Genomic biomarkers of folate status

DNA cytosine methylation. Apart from its role in DNA synthesis, folate plays an important role in DNA metabolism as it is required for the synthesis of methionine and thus SAM, which in turn is required as a methyl donor for the maintenance of cytosine methylation, essential for silencing of genes or structural integrity of specific regions of chromosomes (Fenech, 2012). When SAM is depleted, the maintenance of DNA methylation may become compromised, leading to hypomethylation of cytosine and structural changes in chromatin. Some studies suggest that global DNA methylation status is reduced when folate is deficient (Kim, 2005; Crider et al., 2012), but findings in this regard are somewhat inconsistent.

Uracil misincorporation into DNA. Measurement of uracil content in DNA may also provide a biomarker of folate status, on the basis that adequate amounts of folate as 5,10-methyleneTHF are required by thymidylate synthase to convert deoxyuridine monophosphate (dUMP) to thymidylate monophosphate (dTMP) in pyrimidine synthesis and thus DNA biosynthesis. If 5,10-methyleneTHF is limiting as a substrate of thymidylate synthase, dUMP accumulates and it becomes more probable that uracil is incorporated into DNA instead of thymidine during DNA synthesis (Stover, 2009; Fenech, 2012).

Micronuclei. Micronuclei have the same morphological features as normal nuclei but are much smaller. Excessive uracil incorporation into DNA and hypomethylation of DNA can lead to the formation of micronuclei from chromosome fragments which can be measured in lymphocytes or in erythrocytes (Blount et al., 1997; Fenech, 2012). Observational studies show that micronuclei in lymphocytes or erythrocytes are inversely associated with dietary folate intake and RBC folate (Blount et al., 1997), whilst intervention studies have reported a significant reduction in micronuclei frequency in lymphocytes with folic acid supplementation (Fenech et al., 1997). It is now possible to score micronuclei automatically and reliably using a wide range of image cytometry platforms making this technique amenable to mass screening. Given its sensitivity to folate deficiency, micronuclei measurement, in combination with uracil and DNA methylation measurements, can potentially provide a reliable assessment of genome pathology resulting from deficient folate status (Bailey et al., 2015).

22a.2.4 Assessment of folate status in children

Most studies assessing folate status have focused on adults. Far fewer studies provide folate biomarker data for children, with notable exceptions being published reports using data from population-based surveys conducted in the United States (Pfeiffer et al., 2012) and the United Kingdom (Kerr et al., 2009;. Figure 22a.3) The latter report proposes normal ranges for folate biomarkers for use in clinical pediatric settings

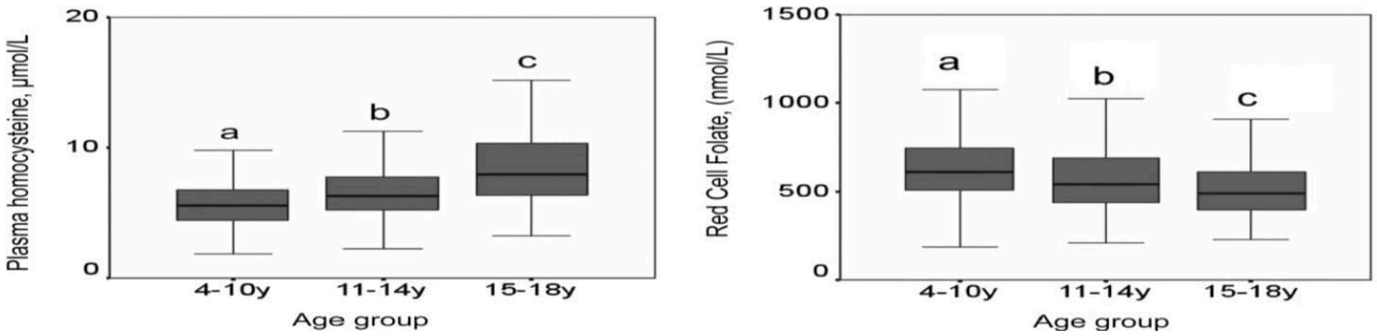


Figure 22a.3 Plasma homocysteine and RBC folate concentrations in a representative sample of British children aged 4–18y. Differences between groups were assessed by using one-factor analysis of covariance (with Tukey's post-hoc test), adjusting for sex, smoking, fortified breakfast cereal consumption, and supplement use. Bars not sharing a common letter differ, $P < 0.05$. Adapted from Kerr et al (2009).

Population-based data in British and American children, both show progressive declines in folate (and corresponding, increases in homocysteine) concentrations with age from childhood to adolescence. Consistent with these reports, are the findings from convenience cohorts of Belgian, Dutch, and Greek children, also showing age-related decreases in folate concentrations in childhood (De Laet et al., 1999; van Beynum et al., 2005; Papandreou et al., 2006). Of note, where dietary intakes were also measured, the data showed that folate intakes generally compared favorably with dietary reference values across all age groups and were not lower in the older children (Papandreou et al., 2006; Kerr et al., 2009; Pfeiffer et al., 2012) The explanation for the decline in folate status biomarkers with age in children, despite no corresponding decline in dietary folate intakes, is not entirely clear but likely reflects the higher folate requirements of older children related to increased metabolic demands for growth from childhood to adolescence (Bailey et al., 2015).

This page titled 21.2: Biomarkers of folate status (Chapter 22a.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

21.3: Conclusions (22a.3)

Folate is required for one-carbon metabolism and thus plays a critical role in essential biological processes, including amino acid metabolism, DNA synthesis and repair and methylation reactions. Low folate status leads to adverse health effects throughout the lifecycle, even if not severe enough to cause megaloblastic anemia, the clinical manifestation of folate deficiency. Thus, accurate assessment of folate status is essential. Serum and RBC folate concentrations are sensitive biomarkers used widely to assess folate status. Serum folate is the earliest indicator of altered folate exposure and reflects recent dietary intake, whilst RBC folate reflects tissue folate stores and is a better indicator of folate intake over the previous 3–4 months when circulating folate is incorporated into the maturating red cells. On the basis that normal homocysteine metabolism requires an adequate supply of folate, the measurement of plasma homocysteine provides a sensitive functional indicator that will be elevated with deficient or low folate status, but it is not a specific folate biomarker as it is influenced by other nutrient (most notably vitamin B_{12} deficiency) and non-nutrient factors including renal function. Emerging genomic biomarkers may provide a reliable assessment of genome pathology resulting from deficient status to folate, to add to the core biomarkers of status, serum folate, RBC folate, and plasma homocysteine.

This page titled 21.3: Conclusions (22a.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

22: VITAMIN B12 (CHAPTER 22B)

Abstract

Dietary deficiencies of vitamin B12 in industrialized countries are relatively rare and usually arise from food-bound malabsorption, particularly in the elderly. Secondary B12 deficiency may also develop from certain disease states or medications that interfere with B12 absorption. In low or middle-income countries, however, dietary vitamin B12 deficiency may occur among children and pregnant and lactating women who exclude or consume low amounts of animal-source foods. Absorption of B12 occurs via passive or active absorption, the latter a complex process during which the bioavailability of dietary B12 in healthy adults is assumed to be about 40-50%. Dietary requirements for B12 vary across countries, with some establishing adequate intakes (AIs) but not requirements per se. There is no Tolerable Upper Intake Level (UL) for vitamin B12 because high intakes have not been associated with any adverse health effects.

There are two major metabolic roles for vitamin B12: (a) synthesis of methionine from homocysteine and (b) conversion of methylmalonyl coenzyme A to succinyl coenzyme A. Vitamin B12 has an important role in cellular metabolism, especially in DNA-synthesis, methylation and mitochondrial metabolism. Deficiency of B12 predominantly affects hematological and neurological systems, although neuropsychiatric symptoms are often the first clinical manifestation. Notably, the full spectrum of clinical features associated with B12 deficiency is much broader, is nonspecific, and may occur as a consequence of other diseases.
Risk of B12 insufficiency can be assessed through measurement of dietary vitamin B12 intake, a biomarker of exposure. Currently, no gold standard exists to define deficiency of vitamin B12. Total B12 concentration in serum is most frequently used, despite its poor sensitivity and specificity. Consequently, a combination of biomarkers is preferred to define and classify B12 status. If subclinical vitamin B12 is suspected, then circulating total vitamin B12 or transcobalamin (TC)-bound B12 (i.e., holo TC) in serum should be measured first, followed by methylmalonic acid (MMA) or homocysteine in serum, both more sensitive functional biomarkers that reflect B12 stores rather than intake. Algorithms based on all or some of these biomarkers measured either sequentially or concurrently are often applied, with concurrent biomarkers combined into a single indicator termed cB12. For all B12 biomarkers, information should be collected on age, gender, ethnicity, pregnancy, folate status, and presence of possible bacterial overgrowth because all these factors impact B12 biomarkers. There is no consensus on the interpretive criteria used to define inadequate B12 status with these B12 biomarkers. For some, statistically defined reference limits are used, whereas for others "true" cutoffs" based on a known relationship between the B12 biomarker and low body stores, functional impairment, or clinical signs of deficiency, are applied.

Often in clinical patients, once B12 deficiency is diagnosed, the next step is to establish whether malabsorption is the cause. In the past, the Schilling test was used, but is rarely conducted today because this test uses radioactive B12. Instead, the CobaSorb test that measures change in serum holoTC concentrations following an oral dose of unlabeled B12 can be used. Under development is the ^{14}C or ^{13}C-labeled absorption test, each of which measures enrichment of ^{14}C or ^{13}C, respectively in blood, urine, or stool following oral ingestion of microbiologically-produced B12 labeled with ^{14}Carbon or ^{13}Carbon.

22.1: Introduction (22b.1)

22.2: Biomarkers of vitamin B12 status (22b.2)

This page titled 22: Vitamin B12 (Chapter 22b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

22.1: Introduction (22b.1)

The term "vitamin B12" is a generic descriptor for all the corrinoids (compounds containing the corrin nucleus) that exhibit the biological activity of cyano-cobalamin. The latter was first isolated in the crystalline form in 1948 (Rickes et al., 1948; Smith, 1948) and the structure, described by Hodgkin et al. (1956), is shown in Figure 22b.1. Here, a cyano-group is linked to a cobalt nucleus. Different forms of cobalamin exist with linkages of methyl, adenosyl or hydroxo groups to cobalt, resulting in methylcobalamin, adenosylcobalamin and hydroxocobalamin, respectively.

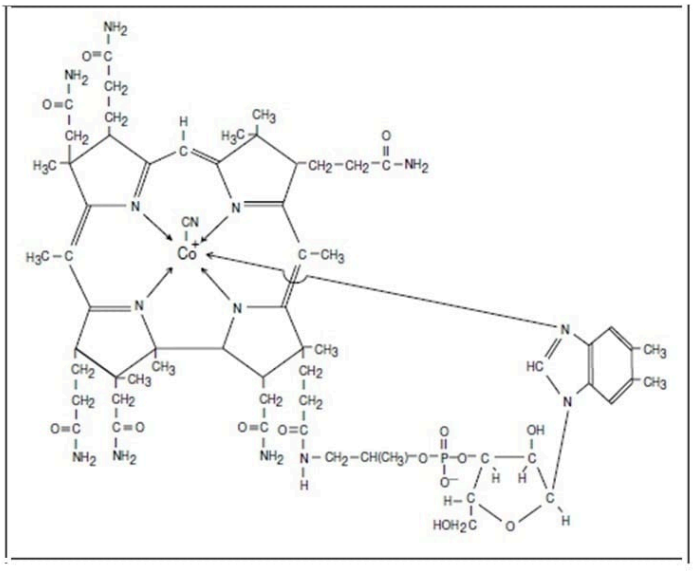


Figure 22b.1: The structure of cyanocobalamin. The molecule is made up of a planar group, consisting of a corrin ring with a cobalt atom at the center and a nucleotide set at right angles to the corrin ring. The nucleotide is made up of the base, 5,6-dimethylbenzimidazole, and a phosphorylated sugar (ribose-3-phosphate).

22b.1.1 Functions of vitamin B12

There are two naturally occurring cobalamin containing co-enzymes: methylcobalamin, the main form in plasma, and 5-deoxy-adenosylcobalamin, found in the liver, most body tissues, and foods. In humans, methylcobalamin functions in the folate-dependent methylation of homocysteine to methionine (Figure 22b.2) whereas 5'-deoxyadenosyl-cobalamin plays a major part in the conversion of L-methylmalonyl-coenzyme A to succinyl-coenzyme A in an isomerization reaction. The latter is a common pathway for the degradation of certain amino acids and odd-chain fatty acids.

Vitamin B12 also has a link with nucleic acid metabolism via its role in the conversion of 5-methyltetrahydrofolate to tetrahydro-folate. Consequently, a deficiency of vitamin B12, like folate, impairs the production of tetrahydrofolate, which is required for thymidine and thus DNA synthesis. Impaired DNA synthesis is responsible for the megaloblastic bone marrow, a characteristic of deficiencies of both vitamin B12 and folic acid. Figure 22b.2.

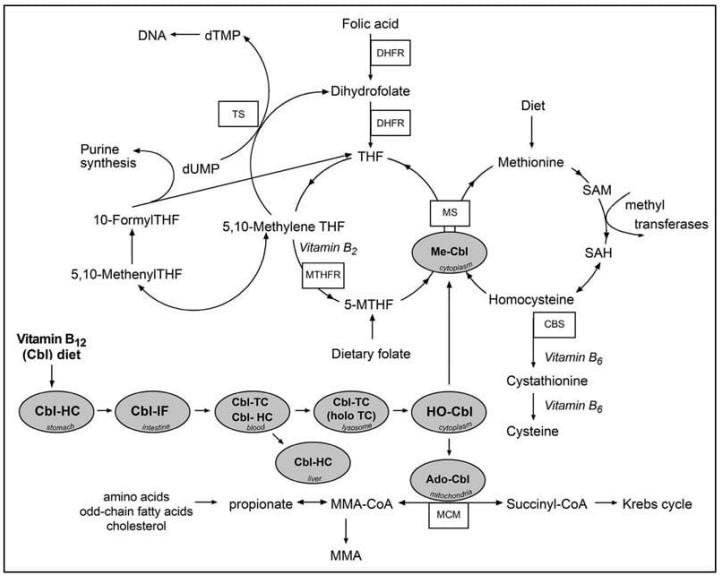


Figure 22b.2 Vitamin B12 absorption and intracellular processing via two enzymatic pathways. In the absence of vitamin B12, 5-MTHF becomes metabolically trapped in this form producing a pseudo folate-deficient state (methyl-trap) and cannot be utilized for regeneration of THF. Cbl: cobalamin; CBS: cystathionine beta-synthase; dTMP: deoxythymidine monophosphate; dUMP: deoxyuridine monophosphate; DHFR: dihydrofolate reductase; HC: haptocorrin; holoTC: holotranscobalamin; HO-Cbl: hydroxocobalamin; IF: intrinsic factor; MS: methionine synthase; Me-Cbl: methylcobalamin; MTHFR: methylene tetrahydrofolate reductase; MMA: methylmalonic acid; MCM: methylmalonyl-CoA mutase; 5-MTHF: 5-methyltetrahydrofolate; SAH: S-Adenosyl homocysteine; SAM: SAdenosyl methionine; THF: tetrahydrofolate; TS: thymidylate synthase; TC: transcobalamin. Redrawn from (Sobczynska-Malefora & Smith, 2022)

22b.1.2 Food sources and dietary intakes

In nature, vitamin B12 is produced almost entirely by bacterial synthesis. Vitamin B12-producing bacteria colonize the large bowel, but humans do not have the ability to absorb bacterial vitamin B12 from their large intestine. Instead, humans are dependent solely on dietary sources of vitamin B12 from: (a) methyl- and 5'deoxyadenosyl forms of vitamin B12 present in animal-source foods, which reach the small intestine (Sobczynska-Malefora & Smith, 2022) or (b) crystalline B12 in fortified foods or dietary supplements. Plants do not contain vitamin B12, except for some dried green and purple seaweeds, which have a symbiotic relation with B12 producing bacteria (Allen et al., 2018). The richest food sources of vitamin B12 include meat (especially liver, followed by kidney and heart), eggs, dairy products, fish and shellfish. Several forms of vitamin B12 exist in foods: meat and fish contain mainly adenosyl-cobalamin (Ado-Cbl) and hydroxocobalamin (OH-Cbl). Organo-cobalamin (OG-Cbl) is found mostly in milk; and methyl-cobalamin (Me-Cbl), Ado-Cbl, and OH-Cbl are found in nearly all dairy products (Sobczynska-Malefora et al., 2021).

Vitamin B12 is fairly stable and normally not destroyed by cooking at normal temperatures, although in an alkaline pH some loss may occur. Bioavailability of vitamin B12 from foods is generally assumed to be 40-50% for healthy adults. However, studies assessing bioavailability of B12 from specific food sources in healthy participants have shown that absorption varies depending on both the specific food product and the magnitude of the dose, ranging from 36%-24% for egg products (dose 0.3-0.94mg, respectively), 83%-52% for lean meat (dose 0.54-5.11mg, respectively), 42%-30% for fish (dose 2.1-13.1mg, respectively) and 49%-4.5% for liver products (dose 0.5-38mg, respectively) (Doets et al., 2013).

Early approaches for studying the bioavailability of B12 involved radio-actively labeled B12. New methods under development for studying the active absorption of B12 from different foods include the use of ^{14}C-labeled B12 (Brito et al., 2018) and the stable

isotope ^{13}C-labeled B12 (Devi et al., 2020). The uptake of supplemental (crystalline) vitamin B12 from B12-fortifed foods and dietary supplements occurs by passive diffusion.

22b.1.3 Absorption and metabolism

For the transport into body cells, vitamin B12 is absorbed by either active absorption or passive diffusion. Only 1-2% of an oral crystalline dose of non-protein bound vitamin B12 — as occurs in fortified foods and supplements — is absorbed by simple passive diffusion throughout the length of the gastrointestinal tract (Doets et al., 2013).

Active absorption of vitamin B12 is a highly complex active transport system, which takes 3-4h to complete from ingestion by mouth to appearance in the circulation. Essential steps include the release of dietary B12 from the food matrix by the activity of hydrochloric acid and gastric protease in the stomach. Here B12 binds to haptocorrin produced by salivary glands. In the upper small intestine where the pH changes to less acidic conditions, haptocorrin is degraded by pancreatic proteases, releasing B12 which subsequently becomes bound to gastric intrinsic factor (IF) to form the IF-B12 complex. Gastric intrinsic factor is a highly specific binding glycoprotein, which is synthesized and secreted by the parietal cells of the gastric mucosa.

Absorption of the IF-B12 complex mainly takes place through receptor sites in the terminal ileum, at an alkaline pH, and in the presence of calcium ions (Allen et al., 2018; Herbert, 1987). However, the capacity of IF-mediated active absorption is rate limited, as the IF-B12 ileal receptors become saturated at B12 intakes of about 1.5-2.5µg from a single meal (Scott, 1997). Reports of isotope studies, including the use of the stable isotope ^{13}C-labeled B12 (Devi et al., 2020), indicate that 50% of an oral dose of 1µg is absorbed, 20% of a 5µg dose, and 5% of a 25µg dose, as shown in Figure 22b.3.

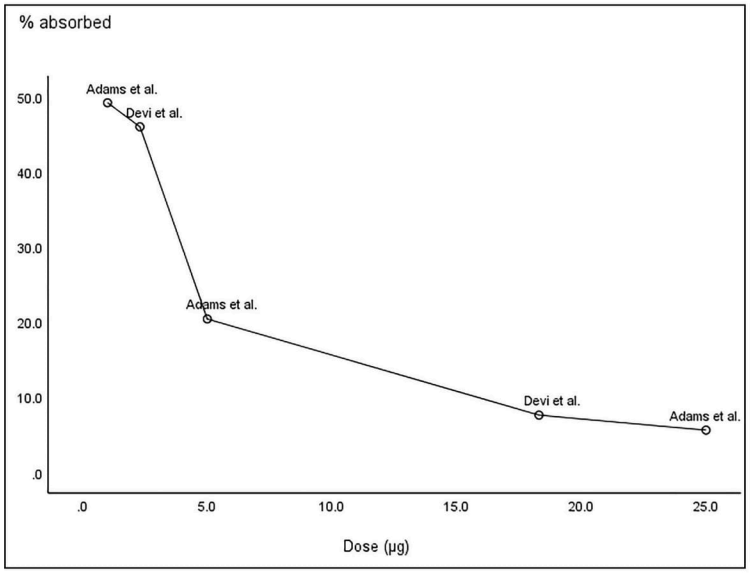


Figure 22b.3 Relationship between the intake and percentage of absorbed B12 from a dose of ^{58}Co-cyanocobalamin (Adams et al., 1971) and ^{13}C-cyanocobalamin (Devi et al., 2018) in 10-12 healthy volunteers. Redrawn from Sobczynska-Malefora et al., 2021.

Once in the blood, free B12 binds to two proteins: transcobalmin, forming holo-transcobalamin (holoTC), and haptocorrin to form holo-haptocorrin (holoHC). Of the two, holoTC has a much faster turnover than the holohaptocorrin complex, and as a result low TC concentrations are considered to be the earliest indicator of failure to absorb vitamin B12 from food. Normally only about 10% of total transcobalamin is bound to B12, whereas in contrast haptocorrin carries about 80% of the B12 in the circulation, despite having no role in the cellular uptake of vitamin B12 (Sobczynska-Malefora & Smith, 2022). Upon delivery of B12 to the tissues by

holo TC, receptor-mediated endocytosis occurs, after which holoTC enters acidic lysosomes where the transcobalamin protein is degraded, the B12 is released, and then converted to its cofactor forms (Allen et al., 2018).

The main storage site for vitamin B12 is the liver; estimates for mean values for total body B12 stores range from 1.1 and 3.9mg (Doets et al., 2013). Loss of vitamin B12 takes place via desquamation of epithelium and through secretion in the bile. Most of the vitamin B12 secreted in the bile, however, is reabsorbed and is thus available for metabolic functions. Losses of B12 in adults approximate 1-3µg/d (about 0.1% of body stores). However, because excretion of vitamin B12 in the stool is proportional to B12 body stores, B12 deficiency develops more slowly in persons han in persons with no intrinsic factor or with malabsorption

22b.1.4 Deficiency of vitamin B12 in humans

The common consequence of B12 deficiency is a cellular deficit in one of the coenzyme forms of B12 which affects cellular metabolism, methylation and mitochondrial metabolism: such deficits may pass from a stage of subclinical deficiency through to severe, clinical forms. The clinical signs of vitamin B12 deficiency occur in the hematopoietic, gastrointestinal, and neurological systems Figure 22b.4.

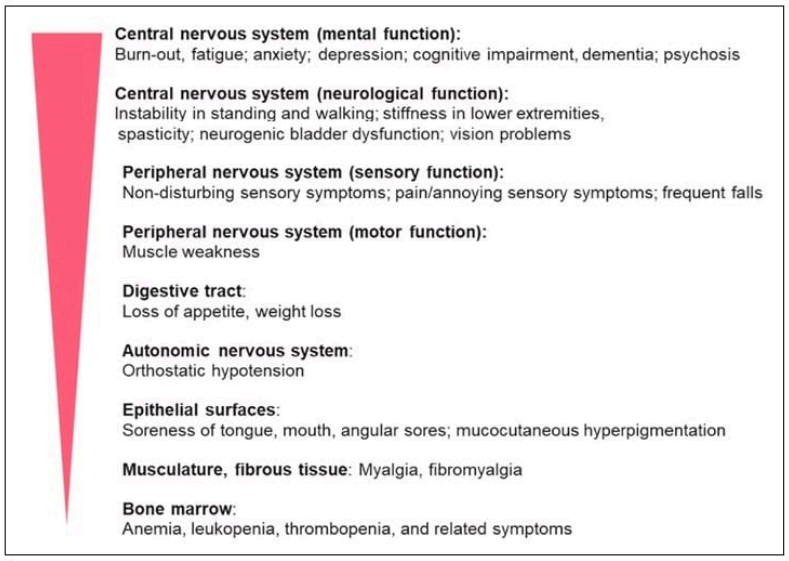


Figure 22b.4 Symptoms of B12 deficiency ordered from the most to the least difficult for being linked to B12 deficiency. Modified from Obeid et al., 2024. Note, although neuropsychiatric symptoms are more frequent than hematological abnormalities, they are less specific, more demanding to recognize, and thus more difficult to link to B12 deficiency

The hematological effects of B12 deficiency are indistinguishable from those of folate deficiency and include megaloblastic anemia, accompanied by the classical anemia symptoms of diminished energy and tolerance to exercise, fatigue, shortness of breath, and palpitations. The megaloblastic changes result from an interference with normal DNA synthesis, arising from lack of 5,10-methylene tetrahydrofolate, as described for folate deficiency.

Gastrointestinal effects of vitamin B12 deficiency include atrophic glossitis, papillary atrophy of the tongue, loss of appetite, flatulence, and constipation.

The neurological complications progress gradually and are variable. They may recede or occur in the absence of anemia (Lindenbaum et al., 1998; 1995). Manifestations include tingling and numb extremities, motor disturbances such as abnormalities of gait, and cognitive changes that range from memory loss and depression to frank dementia, with or without mood changes (Goodman & Salt, 1990). Some of the neurological complications result from defective myelin synthesis and repair, and not all are reversible.

Although vitamin B12 deficiency with these classic hematologic and neurologic signs and symptoms is rare (Hughes-Jones & Wickramasinghe SN, 1996), low or marginal vitamin B12 status (serum B12 200-300pg/mL [148-221pmol/L]) without these

symptoms is more common (i.e., up to 40% in Western populations), especially in individuals with low intakes of vitamin B12-rich foods (NIH, 2024).

Interest in the possible association of B12 deficiency with Alzheimer's disease, vascular dementia, and other neurological disorders has increased with the recognition that homocysteine can be a biomarker for B12 deficiency. In elderly persons, for example, serum total homocysteine concentration is often elevated in those whose folate status is normal but who have a clinical response to B12 treatment (Stabler et al., 1996). Moreover, in several reports cognitive decline and brain atrophy has been slowed by lowering plasma total homocysteine concentrations with B vitamins including B12(Smith et al., 2018). These findings have led to the consensus that elevated plasma total homocysteine is a modifiable risk factor for development of cognitive decline, dementia, and Alzheimer's disease in older persons. Convincing arguments also exist for the role of homocysteine as a biomarker for cardiovascular disease and the utility of B vitamins and folate as a treatment. According to Smith and Refsum (2021), homocysteine values of 10μmol/L are probably safe in adults but for levels 11μmol/L or above there may be a risk for adverse effects, justifying intervention. More research is warranted.

Globally the prevalence of vitamin B12 deficiency varies widely (Allen et al., 2018). Notably in some countries the prevalence exceeds 40% in specific subpopulations (children, young adults, pregnant and lactating women) especially among those who exclude or consume low amounts of animal products in their diets (Allen, 1994). There are also some reports of deficiency symptoms developing among infants born to mothers deficient in vitamin B12 (Allen, 1994).

In most industrialized countries, clinical B12 deficiency is relatively rare, with many of the causes being of ileal or gastric origin (Green & Miller, 2022). Pernicious anemia (affecting 0.1% of the general population, and up to 2-4% of adults over the age of 60y) or malabsorption associated with a variety of other conditions are usually the cause of B12 deficiency (Stabler & Allen, 2004). Pernicious anemia is an autoimmune gastritis with a genetic component that destroys the gastric mucosa. Consequently, with the lack of gastric intrinsic factor, IF-mediated active absorption in the ileum is compromised. Pernicious anemia can also be caused by long-term gastric atrophy or atrophic gastritis which result in loss of gastric acid. The latter hinders the release of B12 from food and can also cause bacterial overgrowth with competition for the vitamin (Allen et al., 2018).

A variety of other disease states may result in the malabsorption of B12 and thus, secondary vitamin B12 deficiency. For example, in patients who have had a total gastrectomy malabsorption of vitamin B12 occurs through a lack of intrinsic factor. Megaloblastic anemia in these patients may not appear until a long period after the total gastrectomy (5-10 y), because of the capacity of the body to reutilize the vitamin. In persons with intestinal lesions such as jejunal diverticulosis or some anatomical abnormalities with bacterial overgrowth (e.g., fistulas and blind-loops), vitamin B12 deficiency may develop because bacteria in the colon competitively utilize all the available vitamin B12 (Cooke et al., 1963). The fish tape worm (Diphyllobothrium latum) also sequesters vitamin B12 and in the past was a well-recognized cause of vitamin B12 deficiency in Finland. Patients with either tropical or nontropical (gluten-sensitive) sprue and regional ileitis may also develop vitamin B12 deficiency, because of alterations in the brush border structure of the ileal mucosa, which contains the receptor for intrinsic factor. The deficiency may also occur in pancreatic insufficiency due to an inability to digest protein-bound vitamin B12.

Some drugs, including proton pump inhibitors (PPIs) and histamine 2 receptor antagonists (H2RAs) are acid-inhibiting, impairing the release of dietary B12 from food proteins, thus producing risk of B12 deficiency (Lam et al., 2013). Metformin treatment of diabetes causes B12 to accumulate in the liver, thus lowering serum B12 concentrations (Allen et al., 2018). In the UK, healthcare professionals are advised to check and monitor serum vitamin B12 levels in patients treated with metformin and who have symptoms suggestive of vitamin B12 deficiency (https://www.gov.uk/drug-safety-updat...d-vitamin-b12- levels-new-advice-for-monitoring-patients-at-risk).

Nitrous oxide (N_2O) is commonly used for sedation and pain relief, but it is also subject to abuse. The gas chemically inactivates B12 through irreversible oxidation of its coenzyme form, methylcobalamin, at the active site of the B12-dependent methionine synthase reaction. Depending on the frequency and duration of exposure and the vitamin B12 status of the individual, B12 deficiency may develop quickly or more slowly (Green et al., 2017).

Some genetic diseases that involve defects of transcobalamin, methylmalonyl CoA mutase, or enzymes in the pathway of cobalamin adenosylation, may also result in vitamin B12 deficiency (IOM, 2000). For more details on vitamin B12 status in health and disease, see Sobczynska-Malefora et al. (2021).

22b.1.5 Recommended intakes of vitamin B12

When setting recommended dietary reference values, expert groups assume the bioavailability of B12 from the diet is about 50% for healthy adults with normal absorption, even though the absorption from specific foods varies markedly as noted earlier. Absorption of crystalline B12 from fortified foods and dietary supplements is higher ranging from 55 to 74%, and unaffected by food-bound malabsorption.

Currently there is some debate about whether the current B12 recommendations are sufficient to achieve optimum vitamin B12 status (Doets et al., 2013). Differences exist across expert groups on both the approaches used, and the recommended B12 levels set. For a summary table comparing the recommendations across countries, see Allen et al. (2018). In the United States and Canada, Dietary Reference Intakes (DRIs) were set to maintain hematological values of serum vitamin B12 within the normal range, whereas the European Food Safety Authority (EFSA) concluded there was insufficient available data to set DRIs (EFSA, 2015). Instead, EFSA used a combination of B12 biomarkers and European adult mean B12 intakes to estimate Adequate Intakes (AIs) for all life-stage and physiological groups.

In the United States and Canada, the Estimated Average Requirement (EAR) for both males and females aged 19 >70y is 2µg/day. The EAR represents the level of vitamin B12 that is estimated to meet the B12 requirements of 50% of the healthy individuals in a particular sex and life-stage group. The corresponding Recommended Dietary Allowance (RDA) for these adults is 2.4µg/day (assuming a CV of 10%), a level designed to cover the needs of 97 to 98 percent of individuals in the group. However, because of the high prevalence of food-bound malabsorption among older adults, those >50 years are encouraged to consume B12 fortified foods (such as fortified ready-to-eat cereals) or supplements to ensure they meet their B12 needs. For details of the DRIs for other life-stage groups, see IOM (2000) and Allen et al. (2018). WHO/FAO (2004) has adopted the DRIs for B12 set by IOM (2000).

In contrast, the AIs set by EFSA (2015) as an intake goal for individuals in all life-stage and physiological groups, are higher than those set by the IOM (2000). and WHO/FAO (2004) and depending on age, range from 1.5 and 4.0µg/day, for adults. For pregnant and lactating women, the AIs are 4.5 and 5.0µg/day, respectively. The AIs for infants and children were calculated by extrapolation from the AI for adults. For more information on the B12 intakes from nine European countries, see EFSA (2015). For more details on the correct use of these differing dietary B12 recommendations, see Chapter 8b: Evaluation of nutrient intakes and diets.

22b.1.6 Effects of high intakes of vitamin B12

There appears to be no evidence that excessive intakes of vitamin B12 from food or supplements causes adverse effects in healthy persons (IOM, 2000). Consequently, both the U.S. Food and Nutrition Board (IOM, 2000) and EFSA (2024). have concluded there are insufficient data to set a Tolerable Upper Intake Level (UL) for vitamin B12.

This page titled 22.1: Introduction (22b.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

22.2: Biomarkers of vitamin B12 status (22b.2)

Biomarkers of vitamin B12 status have been reviewed by the expert international panel as part of the Biomarkers of Nutrition for Development (BOND) project; see Allen et al (2018) for more details. Risk of B12 insufficiency can be assessed through the measurement of dietary intake of vitamin B12, a biomarker of exposure. Quantitative dietary assessment methods such as 24h recalls and food records must be used; for more details see Chapter 3: Measurement of food consumption of individuals. The variation in day-to-day intake of vitamin B12 is often large when vitamin B12-rich foods such as liver are consumed, so multiple days of intake should be recorded on each individual and the observed distribution of intakes adjusted to yield usual intake distribution. This is achieved by removing the variability introduced by the day-to-day variation in B12 intakes within an individual. In this way, the proportion of individuals in a population group at risk of B12 insufficiency can be determined. Details of the adjustment process are given in Chapter 3.

In population groups consuming crystalline B12 in fortified foods and in supplements, consumption is often intermittent necessitating several days of intake or use of a food frequency questionnaire. Details of the B12 content of many dietary supplements are available in the U.S. federal Dietary Supplement Ingredient Database. No correction is made for differences in the bioavailability of crystalline B12 from supplements or fortified foods compared with B12 from food because the absorption differences between these two B12 sources are assumed to be small.

Alternatively, instead of determining quantitative B12 intakes, estimates of the usual consumption of animal products per day, week, or month can be used to assess risk of inadequate intakes of B12 for individuals in households and for population groups; see Chapter 3 for more details. Several studies have reported significant positive associations between animal source food intake and vitamin B12 status (McLean et al., 2007).

A combination of biochemical biomarkers is recommended to assess actual B12 status, together with knowledge of factors such as age, gender, pregnancy, folate status, and bacterial overgrowth, all known to impact some of the B12 biochemical biomarkers. The four priority biochemical biomarkers recommended by the Biomarkers of Nutrition for Development (BOND) project to detect subclinical vitamin B12 deficiency are total serum vitamin B12, serum holotranscobalamin (holo TC), plasma homocysteine, and serum methylmalonic acid; see Allen et al., 2018 for more details. Biochemical assays for each of these biomarkers together with the recommended tests to determine the cause of the vitamin B12 deficiency are outlined below.

The BOND project caution that currently no gold standard exists to define vitamin B12-deficiency using these biochemical biomarkers. If subclinical vitamin B12 deficiency is suspected, then total serum vitamin B12 or serum holoTC is frequently measured first, followed by methylmalonic acid (MMA) or plasma homocysteine (tHcy) concentrations.

22b.2.1 Serum vitamin B12

Of the vitamin B12 in the serum, 20-30% is attached to the transport protein transcobalamin (previously known as transcobalamin II), and the remaining 70-80% is bound and carried by haptocorrin, the latter historically designated as transcobalamin I and transcobalamin III. Hence, total plasma or serum B12 concentration includes B12 (cobalamin) bound to transcobalamin (holoTC) and B12 bound to haptocorrin. Assessment of total B12 in plasma or serum provides information on intake of B12, long-term B12 status of the individual, and liver stores.

Of these two circulating transport proteins, only transcobalamin bound to B12 to form holotranscobalamin (holoTC) in serum is metabolically active, delivering vitamin B12 to receptors on cell membranes. In contrast, because there are no haptocorrin receptors located on most cells, B12 cannot be delivered to extrahepatic tissues by haptocorrin. Moreover, haptocorrin binds not only B12 but also B12 analogs. However, because gastric intrinsic factor does not bind the B12 analogs (only B12), the B12 analogs are not reabsorbed but ultimately excreted in the stool (Renz, 1999).

In early vitamin B12 deficiency, when individuals are in negative balance, the amount of vitamin B12 attached to holoTC falls, but there is no concomitant decline in the total serum vitamin B12 concentrations. The latter often remain normal for weeks or months despite low serum holoTC levels. Serum vitamin B12 concentrations only decline when the percentage saturation of total transcobalamin with vitamin B12 falls below 5%(Herbert, 1987).

Total serum vitamin B12 concentrations are expressed aspmol/L, or aspg/mL in clinical practice with 1.0pg/mL=0.7378pmol/L. Concentrations above the normal level of 221pmol/L are indicative of an adequate B12 status, whereas those between 150 and 221pmol/L, or often below 150pmol/L, are indicative of B12 depletion or deficiency, respectively.

Traditionally, total serum vitamin B12 is often the first-line biochemical test used for routine screening for vitamin B12 deficiency; concentrations reflect both the vitamin B12 intake (Dullemeijer et al., 2013) and body stores. However, to date a single accepted cut-value for serum B12 indicative of deficiency has not been established. Moreover, the serum B12 test has low sensitivity and specificity and may overestimate B12 deficiency when used alone (Mineva et al., 2021). For example, clinical cases of vitamin B12 deficiency have been reported even in persons with low-to-normal total serum vitamin B12 levels (Lindenbaum et al., 1988). The latter may be explained by the presence of a physiologically inactive proportion of serum B12 being bound to the carrier protein haptocorrin, as explained earlier. Many, but not all of these cases, have shown a positive response to treatment with vitamin B12, based on improvements in abnormal metabolite levels and clinical indices (Stabler et al., 1990).

Most available prevalence data on B12 deficiency based on total serum B12 concentrations are focussed on older adults. In reports based on 48 cohorts, 10-19% of adults more than 60 years of age had total serum B12 levels below 148 or 150pmol/l and said to be indicative of B12 deficiency. A high prevalence of low serum / plasma B12 has also been reported among pregnant women (27.5% in 11 cohorts) and children (12.5% in 14 cohorts)(Smith et al., 2018).

Factors affecting total serum vitamin B12

Age influences levels of total serum vitamin B12. Infants and children have higher serum B12 concentrations compared with adults, irrespective of ethnicity, perhaps in part attributed to their higher intake of milk in early childhood and a more efficient hepatic storage of B12 (Abildgaard et al., 2022; Sobczynska-Malefora et al., 2023). During adulthood, serum B12 concentrations tend to decline with age, especially in the elderly (Wahlin et al., 2002), probably arising from the gradual decrease in both gastric acidity and intrinsic factor production that occurs with aging (Asselt et al., 1998; Carmel, 1997). An age-related decrease in the intake of vitamin B12 may be an additional factor (Johnson et al., 2003). However, such a decline in serum B12 among elderly adults is not consistent (Vogiatzoglou et al., 2009; Mineva et al., 2019), possibly due to B12 supplement use in this age group. In the United States NHANES 2011 to 2014 survey, for example, serum vitamin B12 concentrations were reportedly 13% higher in persons aged >70 y compared to those in the 20-39y age group, a trend attributed mostly to B12 supplement usage (Mineva et al., 2019).

Gender may affect serum vitamin B12 concentrations, although results have been inconsistent. In smaller, less well controlled studies levels for men and women have been similar, irrespective of age (de Carvalho et al., 1996; Wahlin et al., 2002). In studies that have controlled for factors known to affect total serum vitamin B12, concentrations were higher in women than in men (Mineva et al., 2019; Metz et al., 1971; Fernandes-Costa et al., 1985), whereas in others no differences between the sexes were found, irrespective of ethnicity (Sobczynska-Malefora et al., 2023).

Pregnancy can lead to a 25-30% fall in total serum vitamin B12 concentrations between 20- and 30-weeks of gestation. The decline is due to decreased synthesis of haptocorrin and hemodilution. Holotranscobalamin (holoTC) concentrations, however, remain unchanged during pregnancy so the decline is unlikely to reflect B12 depletion. Indeed, by 14 weeks postpartum, total serum B12 concentrations are no longer low (Allen et al., 2018).

Lactation is accompanied by a 40% increase in serum B12 at 6 weeks post-partum, remaining at this elevated level while the mother is lactating. A corresponding moderate increase in the functional metabolic biomarkers — plasma methymalonic acid and homocysteine — suggests that some depletion of maternal intracellular vitamin B12 has occurred even though serum B12 concentrations are high (Varsi et al., 2018).

Ethnicity affects serum/plasma B12 concentrations. People of African origin have higher serum B12 concentrations than other ethnic groups. For example, in the US NHANES 2011-2014 survey, non-hispanic black adults had higher B12 concentrations than non-hispanic white adults, even after controlling for age group and B12 supplement use (Mineva et al., 2019). This ethnic difference is attributed in part to a higher concentration of B12 binding proteins, most notably the transport protein haptocorrin. However, because B12 bound to haptocorrin is not metabolically active, use of only total serum B12 may result in Black patients being incorrectly identified as having a high B12 status. Clearly, failure to apply ethnic-related serum B12 reference intervals will lead to underdiagnosis of B12 deficiency and any associated comorbidities (Sobczynska-Malefora et al., 2023).

Folate status, both deficiency and excess, might affect total serum vitamin B12 levels. Folate deficiency may result in moderately

low total serum vitamin B12 concentrations because both folate as a cosubstrate (methyl donor) and vitamin B12 (as a coenzyme) are required in the remethylation metabolic pathway of homocysteine (see Figure 22b.2) (Klee, 2000).

The mechanism for the interaction between excess folate and B12 is uncertain and warrants further investigation. One proposed hypothesis, as yet untested, is that high-dose folic acid supplements cause depletion of the transport protein holotranscobalamin in serum, thus exacerbating B12 deficiency. Older adults with both low total serum B12 concentrations and increased levels of serum folate appear to have a higher risk of anemia and dementia, based on a review of the clinical and epidemiological evidence. The high serum folate levels are presumed to arise from exposure to high intakes of folic acid, the synthetic form of folate used in fortified foods and supplements (Miller et al., 2024).

Bacterial overgrowth of the small intestine may lead to low serum B12 concentrations and vitamin B12 deficiency because the bacteria utilize vitamin B12 for their own metabolism so B12 cannot be absorbed by the body.

Fish tapeworm also leads to B12 malabsorption because the tapeworms consume B12 for their own needs, so serum B12 concentrations fall.

Chronic infection with *Helicobacter pylori*, often accompanied by intestinal bacterial overgrowth, may result in low serum B12 concentrations. The latter arise from malabsorption from food-bound B12 induced by the hypochlorhydria associated with atrophic gastritis caused by the *H pylori* jnfection of the stomach (Sobczynska-Malefora et al., 2021).

Lack of gastric intrinsic factor caused by pernicious anemia or gastrectomy leads to low total serum vitamin B12 concentrations. Intrinsic factor (IF), a glycoprotein synthesized by the gastric parietal cells, is required for the absorption of vitamin B12 in the small intestine. In some cases of pernicious anemia, parietal cell auto-antibodies lead to the destruction of IF-producing parietal cells in the stomach resulting in B12 malabsorption, and thus deficiency. The incidence of pernicious anemia increases with age, and is especially common among persons of European or African descent (Stabler & Allen, 2004). Parietal cell autoantibodies are also found in other autoimmune conditions such as Graves disease, hypothyroidism, and Addison's disease (Sobczynska-Malefora et al., 2021).

Other conditions that produce elevated haptocorrin levels increase total serum B12. Examples of such conditions include myeloproliferative disorders (e.g., chronic myelogenous leukemia, polycythemiavera), severe liver diseases, and alcoholism. In these conditions, although total serum vitamin B12 concentrations may appear elevated or apparently normal, tissue vitamin B12 concentrations will be low (Arendt & Nexo, 2012).

Some disease states are associated with low total serum vitamin B12 concentrations because of an inability to absorb or digest protein-bound vitamin B12. Disease states associated with B12 malabsorption due to villous atrophy or mucosal impairment include ileal Crohn's or Celiac disease, respectively. In disease states such as atrophic gastritis and pancreatic insufficiency, low serum B12 concentrations arise because digestion of protein-bound B12 is compromised. Other intestinal diseases such as tropical sprue, intestinal lymphoma, amyloidosis, and short bowel syndrome have also been associated with low serum B12 levels and B12 depletion.

Surgery involving partial or complete removal of any part of the gastrointestinal tract, including some bariatric surgery procedures, will reduce food-bound B12 absorption and lower serum B12 concentrations.

Inborn errors of intracellular B12 metabolism can result in low total serum B12 levels through their effects on absorption, transport, or metabolism of B12. The genetic defects range from rare gene deletion and mutations to single nucleotide polymorphisms with more mild and subtle effects. In congenital pernicious anemia, low serum total B12 concentrations arise from B12 malabsorption caused by mutations in a gene (GIF,CBLIF) and defects in the synthesis of intrinsic factor. The rare autosomal recessive disorder Immerslund-Grasbeck (IGS) disease is caused by mutations in either the gene Amnion Associated Transmembrane Protein (AMN) or Cubilin (CUBN) gene responsible for the synthesis of cubam receptors (Sobczynska-Malefora et al., 2021).

Genetically predetermined deficiencies of the two B12 transport proteins, congenital transcobalamin deficiency (gene locus 22q11.2-qter) or haptocorrin deficiency (gene locus 11q11-q12), can be associated with total serum B12 values that may be normal, or falsely low, respectively (Amos et al., 1994). Individuals with defects of methylmalonyl CoA mutase or polymorphisms occurring in the genes that code for enzymes in the cobalamin adenosylation pathway, may also have low serum vitamin B12 levels (Kano et al., 1985). Eight genetic defects in B12 metabolism have been identified all of which result in a failure to utilize B12 by the target cells; for more details see Sobczynska-Malefora et al. (2021).

Restricted dietary intake of B12 is associated with veganism and to a lesser extent vegetarianism, dietary patterns that avoid all

or certain animal-source foods, respectively. Vegans have low total serum B12 concentrations unless crystalline B12 from fortified foods or supplements are consumed. To date, how long it takes for vitamin B12 deficiency to develop when individuals adhere exclusively to a fully plant-based diet is unknown.

There have been several reports of low plasma vitamin B12 levels among Asian Indians in Great Britain, the United States, and the Indian Subcontinent, many of whom are likely to practice vegetarianism so intake of animal-source foods is low (Antony, 2003; Antony, 2001; Carmel et al., 2002). Several investigators have reported an inverse relationship between intake of animal-source foods and total serum vitamin B12 levels. In a study in Germany of omnivores and vegetarians, none of whom were taking vitamin supplements, those consuming omnivorous diets had higher serum B12 concentrations than the lacto- or lacto-ovo vegetarians both of whom avoided meat, poultry, and fish, who in turn had higher levels than the vegans (Herrmann et al., 2003). Similar trends in serum B12 concentrations have been reported among British male omnivores, vegetarians, and vegans in the EPIC Oxford cohort study (Gilsing et al., 2010).

Dietary intakes of omnivores are not generally strongly related to serum vitamin B12 levels; low correlations in daily B12 intakes have been linked to the large size of liver vitamin B12 stores (2-3mg) (Doets et al., 2013). A meta-analysis of randomized controlled trials (RCTs) and observational studies investigated the dose-response relationship between vitamin B12 intake and serum or plasma vitamin B12 concentrations. Doubling vitamin B12 intake was estimated to increase the vitamin B12 concentration in serum or plasma by 11% (95% CI: 9.4%, 12.5%) (Dullemeijer et al., 2013). Several cross-sectional studies among healthy adults with normal absorption of B12 show that serum B12 concentrations reach a maximum plateau and stabilize at B12 intakes of 4-7µg/d (Bor et al., 2010). Mode of infant feeding affects total serum B12 levels; infant formula has a higher B12 content than human milk (Greibe et al., 2016).

Vitamin B12 supplement use increases serum B12 concentrations. In the US NHANES 2011-2014 survey, serum B12 concentrations for adults were about 40% higher in crystalline B12 supplement users than nonusers overall (i.e., for age groups 20-39, 40-59, 60-69, and >70 years), with the greatest increase among older persons. In non- supplement users, however, serum B12 concentrations were similar across all four age groups (Mineva et al., 2019).

Medications such as histamine 2 receptor antagonists and proton pump inhibitors (PPIs) (lansoprazole, omeprazole, and esomeprazole) that suppress gastric acid may lower total serum B12 levels by impairing the release of B12 from food (Damodharan et al., 2021; Mumtaz et al., 2022). Metformin, used as treatment for type 2 diabetes, induces B12 malabsorption and has also been associated with low total serum B12 concentrations(Al-Fawaeir & Al-Odat, 2022). However, the low serum B12 levels may not reflect true B12 deficiency in view of the finding in animal studies that metformin increases liver accumulation of vitamin B12 (Greibe et al., 2013; Obeid et al., 2024).

Storage of blood samples prior to separation of the serum affects B12 concentrations if separation is delayed. Such delays can lead to an initial rise in total serum B12 concentrations, followed by a significant drop after 1 to 3 days (Allen et al., 2018). Hence to ensure accurate measurements, it is important to separate the blood samples and freeze the serum as soon as possible after collection.

Analytical method may affect the reported serum vitamin B12 concentrations. Some of the nonfunctional analogs of vitamin B12, as well as the cobalamins, were measured in earlier radio-assays. This led to overestimates of actual serum B12 concentrations. This problem has now been overcome by using purified intrinsic factor which does not react with vitamin B12 analogs as the binder in the immune-enzymatic assays (Klee, 2000). However, in patients with pernicious anaemia these assays should not be used because the intrinsic factor antibodies in the patient's serum interferes with the binding of B12 to the intrinsic factor used in the assay, generating spurious results.

For the analytical platform used it is essential to adopt reference ranges that are compatible with the chosen platform (Harrington, 2017). A standard reference material for serum B12 has been developed by the US National Institute for Standards and Technology (NIST) (SRM 3951). This SRM consists of three sera with differing vitamin B12 concentrations: 100pg/mL; 200pg/mL; 450pg/mL.

Interpretive criteria

Interpretation of serum B12 concentrations deemed as adequate or inadequate is challenging. The normal levels for serum

vitamin B12 concentrations in healthy persons are affected by many factors unrelated to vitamin B12 status, some of which are not considered when interpreting the results. These factors may include age, ethnicity, pregnancy, use of certain medications, individual genetic variation, the assay method and any measurement errors associated with its use. All these factors may affect the sensitivity and specificity of serum B12 as a biomarker of B12 status and result in false-positive or false-negative classification of vitamin B12 deficiency.

Traditionally, adequate concentrations of biomarkers such as serum B12 are defined by a reference interval which covers 95% of the results obtained in a healthy, non-diseased population. In practice, however, it is difficult to compile a “true” healthy reference and few exist for serum B12.

Two methods are used to define whether a biomarker is deemed “inadequate”, both of which have been employed for serum B12 concentrations. In some studies, a serum B12 concentration is considered inadequate when it falls below a statistically defined lower reference limit, often based on the 2.5th percentile value from an “apparently healthy” reference distribution compiled within a country. Hence, values for serum B12 that represent the 2.5th percentile will vary, depending on the local reference distribution applied.

Alternatively, the diagnostic criteria used to identify inadequate B12 status have been based on a known relationship between the serum B12 biomarker and low body stores, functional impairment, or clinical signs of deficiency. In such cases, the term “cutoff” should be used. In practice, the term “cutoff” is often used when the value for inadequate B12 has been based on a statistically defined reference limit. Again, cutoffs for serum B12 designated in this way can also vary, depending on both the local conditions and adverse outcome applied. Frequently, the cutoffs for serum B12 indicative of inadequacy have been set based on the relationship between serum B12 and serum methylmalonic acid. The latter is a functional biomarker considered to be relatively specific and sensitive biomarker of B12 status. For more discussion on the Evaluation of nutritional assessment indices, see Chapter 1: Introduction Section 1.6.

Clearly, the reference limit or cutoff applied to designate inadequate vitamin B12 status may vary, leading to differences in the literature on reports for the prevalence estimates for vitamin B12 deficiency. Reports on the various methods used to interpret serum B12 concentrations as adequate or inadequate are outlined below.

The United Kingdom compiled the mean, median, and critical lower and upper percentiles for serum B12 concentrations by sex and age for children, young people aged 4 to 18y, and people aged >65y based on data from the U.K. National Diet and Nutrition Surveys (Gregory et al., 1995; Gregory et al., 2000; Finch et al., 1998). More recently, Kerr et al. (2009) have proposed age-specific 5th and 95th percentile reference ranges for serum B12 based on data from a representative sample of British children aged 4 to 18y from the National Diet and Nutritional Survey. Serum B12 concentrations decreased significantly with age, with values for those consuming fortified breakfast cereals being significantly higher than non-consumers, a trend that was independent of age and gender. Up to aged 14 years, the 5th and 95th percentiles of serum B12 concentrations were 242-749pmol/L for children aged 4-10 years (n = 317) and 172-641pmol/L for children aged 10-14 years (n = 263). For children from aged 15-18 years, sex specific 5th and 95th percentiles were compiled. These values were 139-452pmol/L for boys (n = 113) and 108-502pmol/L for girls (n = 132) (Kerr et al., 2009). Serum vitamin B12 assays in this survey were conducted on a semiautomated analyzer using microparticle enzyme immunoassay (MEIA) technology (Kerr et al., 2009).

In the United States, 95% reference intervals for serum B12 concentrations for four age groups have been reported based on data from apparently healthy adults aged 20 > 70 years from NHANES 2011 to 2014 (Mineva et al., 2019). These are shown in Table 22b.1.

Table 22b.1 Serum vitamin B12 medians, central 95% reference intervals, by age group in US persons ≥20y of age, NHANES 2011-2 immunoassay. Data from Mineva et al. (2019).

Age group	Serum vitamin B12, Median (95% CI,pmol/L)	Serum vitamin B12, 5t
All	378 (372, 384)	158-1140
20-39	373 (364, 381)	167-840
40-59	365 (357, 378)	160-1150

Age group	Serum vitamin B12, Median (95% CI,pmol/L)	Serum vitamin B12, 5t
60-69	398 (375, 420)	152-1530
≥70	416 (401, 428)	137-1720

Pregnant and lactating women were excluded from the data set. It is of interest that age, race/Hispanic origin, and vitamin B12 supplement use were significantly associated with serum B12 concentrations in this survey. Non-Hispanic white persons had lower serum B12 concentrations than non-Hispanic black persons. In this study, a “true cutoff” of >300pmol/L was used to define adequate vitamin B12 status (i.e., repletion) based on earlier research using serum methylmalonic acid as a functional biomarker (see Section 22b.2.4).

Abildgaard et al. (2022) established 95% age-adjusted reference intervals for plasma B12 for children, adults, and elderly individuals in a Danish population. Blood samples from healthy individuals were collected and analyzed and routine clinical plasma B12 and methylmalonic acid results extracted to establish reference intervals. The 95% reference intervals for plasma B12 based on blood samples and routine patient data from birth to aged >65 years categorized into five age groups are shown in Table 22b.2.

Table 22b.2 Proposed 95% Reference Intervals (RI) for plasma B12 based on blood samples and patient data from a Danish popula

Age group (y)	95% RI, B12 (pmol/L)	% of B12 patient samples below/above 95%
0 - <1	180-1400	3.3%/0.0%
1 - <12	260-1200	5.5%/2.1%
12 - <18	200-800	2.8%/3.4%
18 - <65	200-600	3.0%/12.1%
≥65	200-600	2.2%/19.2%

Total plasma B12 was measured with an automated competitive chemiluminescent immunoassay. The highest plasma B12 concentrations were found in infants, with levels gradually decreasing with age. A reference interval of 200-600pmol/L was established for all adults, with a lower plasma B12 reference limit of 200pmol/L, irrespective of age. The authors emphasize that when plasma B12 is used alone to screen for B12 deficiency in surveys which include children and adolescents, then age-dependent reference limits should be applied.

Ethnicity has also been shown to be a factor that should be considered when interpreting serum or plasma B12 concentrations (O’Logbon et al. 2022). Dietary habits alone do not appear to explain these ethnic differences. Instead, higher concentrations of B12 binding proteins, most notably haptocorrin, coupled with higher B12-binding capacities in persons of Black ethnic origin, may be responsible for the higher serum B12 levels observed in Black persons compared to their White counterparts.

In response to this concern, Sobczynska-Malefora et al. (2023) have compiled age- and ethnicity-related reference intervals for serum vitamin B12 for a UK population. Unlike the traditional approach whereby data are obtained from individuals deemed “healthy”, these investigators applied an indirect method designed to overcome the ethical and resourcing difficulties experienced when sampling apparently healthy individuals. They accessed large datasets of laboratory results from an ethnically diverse South-east London patient population attending primary care clinics. All laboratories used an automated Chemiluminescent Microparticle Assay (CMIS) with microparticles coated with intrinsic factor for the serum B12 assay. Pregnant women were not excluded.

These laboratory data were categorized into five ethnic groups: Asian or British Asian (Asian), Black or British Black (Black), White, Mixed, or Other ethnic groups. Reference intervals were established for children for four age groups from birth to 13y. For children from age 14y and adults, reference intervals are presented for eight age groups and by two ethic groups (Black and Asian/White)

as shown in Table 22b.3.

Table 22b.3 Age and ethnicity related reference intervals (RIs) with confidence intervals (CIs) (pmol/L) determined by indirect calculations using mined patient test results. Data from Sobczynska-Malefora et al., 2023. The values are shown by age for children up to the age of 13 and for subjects >13y - 49y by age and ethnicity. Data for subjects >50y are given in Sobczynska-Malefora et al. (2023).

Age groups (y)	0-1	2-5	6-9	10-13
Data size (N)	105	474	492	661
95 % CI Lower Limit of RI	142–178	246–308	221–270	168–207
RIs (pmol/L)	159–1025	276–1102	245–798	187–643
95 % CI Upper Limit of RI	971–1081	1042–1165	750–848	593–697
Age groups (y)	14–17	18–29	30–39	40-49
Data size Black (N)	267	1038	1534	1983
95 % CI Lower Limit of RI Black	152–191	146–175	150–186	161–198
RIs Black (pmol/L)	171–639	160–632	167–730	178–807
% Upper Limit of RI Black	585–697	579–691	665–801	739–882
Data size Asian/White (N)	297	2611	3614	3390
95 % CI Lower Limit of RI Asian/White	135–165	119–148	126–161	127–160
RIs Asian/White (pmol/L)	149–456	133–458	143–506	142–501
95 % CI Upper Limit of RI Asian/White	420–494	416–505	460–555	458–548

In this study, the lower reference limits for serum B12 concentrations chosen to designate B12 deficiency for Black and Asian White people aged >13 years were: 166pmol/L and 134pmol/L, respectively. The authors acknowledge that these reference limits require validation using functional B12 biomarkers such as serum methylmalonic acid and homocysteine.

As expected, the Black population had higher serum B12 concentrations than the White population. However, surprisingly there were no differences in serum B12 levels between Asian and White individuals for most age groups in these UK adults. This finding may be associated with the treatment practice in the UK of recommending B12 supplements for Asians; lower serum B12 concentrations in Asians compared to other ethnicities have been reported elsewhere (Quay et al., 2015; Devi et al., 2018). Children in all ethnic groups also had higher serum B12 concentrations compared with adults, perhaps in part attributed to their higher intake of milk in early childhood and a more efficient hepatic storage of B12. Clearly, use of these age- and ethnicity-related reference intervals will assist in preventing under reporting of B12 deficiency in both children and the Black population in the UK.

The BOND project has set cutoffs for serum B12 based on the relationship between serum B12 and serum methylmalonic acid, the latter a relatively specific and sensitive functional biomarker of B12 status (Section 22b.2.4). Recognising the importance of establishing normal levels for serum B12 during infancy and early childhood, the BOND group provide tentative age-specific ranges for adequacy from birth up to age 24 months based on data from healthy breastfeeding Norwegian infants and children. (Mineva et al., 2019, Table 22b.4).

Table 22b.4 Cutoffs for biomarkers of B12. Values are ranges. Data from (Allen et al., 2018).

	Serum B12 (pmol/L)		Serum holotrans-cobalamin (pmol/L)		Serum methyl-malonic acid (nmol/L)	
	Deficient	Adequate/ Normal	Deficient	Adequate/ Normal	Deficient	Adequate/ Normal
Newborn cord blood		120-690		33-240		170-500

	Serum B12 (pmol/L)		Serum holotrans-cobalamin (pmol/L)		Serum methyl-malonic acid (nmol/L)	
	Deficient	Adequate/ Normal	Deficient	Adequate/ Normal	Deficient	Adequate/ Normal
6 mo		121-520		12-90		140-220
12 mo		165-580		19-100		120-830
24 mo		183-260		29-110		120-300
Adults	Severe <75 Deficient 75 - <150 Depleted 150-221	>221	<35-40	40-150 or 40-200	>376 or >271	

For adults, a single cutoff for adequacy represented by the upper limit (i.e., >221pmol/L) of the range said to be indicative of B12 depletion (i.e., 150-221pmol/L) is also presented. In addition, the BOND group also present two serum B12 cutoffs said to be indicative of severe (<75pmol/L) and very severe (75-<150pmol/L) B12 deficiency.

Clearly, although serum B12 is the most frequently used biomarker of B12 status, the marked variation in the literature for the reference limits and cutoffs said to be indicative of inadequate B12 status, make the diagnosis of B12 deficiency and the provision of prevalence estimates based on serum B12 alone challenging. For example, depending on the assay, serum B12 cutoffs for deficiency for adults may range from 120 to 180pmol/L, with 148pmol/L based on 3-SDs from the mean of an adult reference range, said to be the most common cutoff for frank deficiency (IOM, 2000; Yetley et al., 2011).

Moderately low serum B12 concentrations between 148 or 150 and 221pmol/L or 260pmol/L are often considered indicative of depleted or subclinical B12 deficiency. Subclinical mild B12 deficiency is much more common and generally asymptomatic with no hematological or neurological manifestations (Obeid et al., 2024). Such levels are difficult to interpret because they can also occur in association with megaloblastic anemia produced by folate deficiency and in iron deficiency (Amos et al., 1994; Layrisse et al., 1959). Even with very low serum B12 concentrations (i.e., below 75pmol/L), clinically diagnosable symptoms of B12 deficiency are said to be apparent in only about 50% of people (Stabler et al., 1990).

Even with total serum B12 concentrations considered to be within the normal range, neurological or in some cases hematological symptoms related to B12 deficiency have been reported. This discrepancy may arise in part because the majority of B12 in serum is bound to the transport protein haptocorrin which is not physiologically active and thus unavailable for B12-dependent enzymatic reactions in cells (Green et al., 2017). In addition, factors unrelated to vitamin B12 status, and noted earlier, may have contributed to the seemingly normal or sometimes high levels of serum B12.

In an effort to resolve the discrepancies in the serum B12 cutoffs indicative of inadequacy, Bailey et al. (2013) used statistical modeling to identify a single change point at which the relation between plasma methylmalonic acid and serum B12 changes slope to differentiate between inadequate and adequate B12 status. However, Instead of a single change point, they reported three slopes resulting in two change points and three subgroups. The data used were from the earlier US NHANES 1999-2004 surveys. The first group considered at high risk for severe B12 deficiency had serum B12 <126pmol/L and the highest plasma methylmalonic acid (281nmol/L). The second group deemed likely to have adequate B12 status had a serum B12 >287pmol/L and a median methylmalonic acid concentration of 120nmol/L, while the third group was classified as indeterminate and difficult to interpret with an intermediate serum B12 (126-287pmol/L).

Clearly, more research is warranted to establish appropriate cut-offs for interpreting serum B12 concentrations. In response to the concerns noted above that reduce the sensitivity and specificity of serum B12 and confound the diagnosis of B12 deficiency, many investigators now recommend measuring one functional B12 biomarker (serum methylmalonic acid or serum homocysteine) if serum B12 concentrations appear inadequate to ensure a more accurate assessment of B12 status. These additional tests are more sensitive and specific biomarkers of functional B12 deficiency and both may be elevated in persons with serum vitamin B12 levels even in the low to normal range (100-300pmol/L) (Nexø et al., 1994; Stabler et al., 1996). They are described here in Section 22b.2.4 for methylmalonic acid, and in the chapter on folate (Chapter 22a. Section 22a.2.2) for homocysteine.

Measurement of serum vitamin B12

Initially serum B12 was measured by microbiological methods, using bacteria with relatively specific requirements for vitamin B12 for growth. Examples include both *Lactobacillus delbrueckii subsp. lactis* (ATCC 4797), also known as *L. leichmannii,* or *Euglena gracilis* (Herbert et al., 1984). The microbiological assay has several limitations. For example, bacteriostatic substances in the blood, including antibiotics or cancer chemotherapeutic agents, inhibit the growth of the microorganism and interfere with the assay, producing misleadingly low serum vitamin B12 concentrations. The main disadvantage of the microbiological method, however, is its low specificity as it measures a variety of cobalamin analogs that are not necessarily biologically active (Sobczynska-Malefora et al., 2021).

Competitive-binding assays that use radioisotope dilution methods for detection are simpler and less time consuming than the microbiological assays. Many of the available competitive-binding assays methods have been automated (Chan, 1996). and unlike the microbiological assays, they are not affected by antibiotics or cancer chemotherapeutic agents. The coefficient of variation among the six laboratories for serum vitamin B12 analyzed by automated competitive-binding assays has been reported to vary from 4.4% to 10.0% (Klee, 2000). However, the method requires the addition of radioactive cyanocobalamin to compete with vitamin B12 in the serum for the binding sites on an added cobalamin-binding protein. Purified hog intrinsic factor is now often used as the binding protein. Alkaline conditions are used to disrupt the vitamin B12 from the binding proteins, after which the vitamin B12 is generally converted to cyanocobalamin by potassium cyanide, prior to measurement (Klee, 2000). However, the assay is falling into disuse with possible discontinuation of the kits (Allen et al., 2018).

More recently, clinical laboratories are using high throughput automated competitive binding chemiluminescence assays (CBLA), using purified intrinsic factor as a reagent to measure total vitamin B12 after its release from endogenous binding proteins. For this method both samples and comparators should be protected from light during collection and separation. Caution should be observed when this method is used to interpret results in patients with pernicious anemia. In these patients some of the competitive binding chemiluminescence assays can be influenced by the presence of interfering anti-IF antibodies, thereby providing spuriously elevated serum cobalamin concentrations, as noted earlier.

Serum B12 can also be determined using B12-antibodies and B12 enzyme-linked immunosorbent assay (ELISA) kits. Various kits are available. Although these could potentially provide an alternative to IF-based assays, they have not been completely verified (Sobczynska-Malefora et al., 2021).

Both serum and plasma samples in which EDTA has been used as the anticoagulant can be used for all the B12 analytical methods. However, use of lithium heparin should be avoided, as this anticoagulant may produce gelatinous serum and elevated B12 concentrations. Blood samples for analysis of serum B12 should be protected from light during collection, and whole blood separated as soon as possible as delays in centrifugation and separation can lead to an initial rise in serum B12 concentrations (Allen et al, 2018).

22b.2.2 Serum holotranscobalamin

Serum holotranscobalamin (holoTC) is the physiologically active metabolite of vitamin B12 that delivers the vitamin to all DNA synthesizing cells, as discussed earlier. Only 20-30% of the total vitamin B12 in serum is bound to the transcobalamin protein. Holo-TC accounts for 5-20% of total transcobalamin which is made in the ileal enterocytes from intracellularly synthesized transcobalamin and absorbed vitamin B12.

Serum holoTC, like serum B12, reflects a broad range of intakes and status (Yetley et al., 2011), but is more sensitive and specific than serum B12. For example, because holoTC has a short half-life, concentrations quickly fall below normal after vitamin B12 absorption ceases. Hence, low holoTC concentrations in serum are often considered the earliest indicator of negative vitamin B12 balance. With increasing intakes of vitamin B12, concentrations of holoTC rise continuously at least until vitamin B12 saturates the transport proteins. The holoTC assay is unaffected by interference from high-titre intrinsic factor antibody levels. Moreover, serum holoTC is not subject to the 25-30% fall in pregnancy that is bserved with serum B12. At postpartum, serum B12 rises substantially. Instead, serum holoTC increases during pregnancy, a trend that continues to 6 weeks post-partum. Thus, in pregnancy serum holoTC offers a diagnostic advantage over total serum B12 (Varsi et al., 2018). Serum holoTC is also a useful test for identifying patients who may suffer from suspected deficiencies in the two vitamin B12 binding proteins, transcobalamin and haptocorrin.

Serum holoTC is a component of the CobaSorb test now used as a surrogate for the dual isotope Schilling test (Nexo & Hoffman-Lucke, 2011) (see Section 22b.2.7). In the past the Schilling test was the gold standard method to evaluate the functional capacity of the ileal IF-B12 receptor and required the use of cobalt-labeled B12. However, the Schilling test is no longer available worldwide due to safety factors related to use of radioactive B12 and is now considered obsolete.

Factors affecting serum holotranscobalamin

Age, sex, and race and their effects on serum holoTC concentrations appear limited. Only a few reports are available with inconsistent results (Nexo & Hoffman-Lucke, 2011). Hence, currently data are insufficient to take into account the slight differences in serum holoTC concentrations by age, sex, and possibly race for adults.

Pregnancy does not affect serum holoTC, levels increasing from 18 weeks gestation to 6 weeks postpartum. This trend contrasts with the decline in serum B12 observed during pregnancy (Varsi et al., 2018). The latter results from a decrease in the synthesis of haptocorrin and possibly hemodilution. Consequently, assay of serum holoTC rather than serum B12 is preferable in studies of pregnant women.

Inherited defects such as transcobalamin deficiency is associated with unmeasurable levels of holoTC in serum. This genetic defect is associated with severe B12 deficiency because lack of the trancobalamin protein prevents B12 binding and formation of holoTC. Early treatment, however, can lead to a good clinical outcome.

Presence of the rare minor allele re35838-82 (p.R215W) in the transcobalamin gene is also associated with unmeasurable levels of holoTC in serum. In this defect, there is proportionately more B12 bound to serum haptocorrin, the physiologically inactive B12 transport protein. This genetic defect is more common in South Asians and those of African origin, and unlike transcobalmin deficiency, is associated with other B12 biomarkers in the normal range and no clinical symptoms.

Transcobalamin receptor (TCblR/CD320) polymorphisms may also impact on serum holoTC concentrations. In such cases, proportionately more B12 is bound to holoTC leading to elevated serum holoTC levels. Therefore, in such rare conditions (present in 5% of older adults), holoTC might not be a marker of "true" intracellular B12 (Sobczynska-Malefora et al., 2021).

Disease conditions affecting serum holoTC levels include renal patients with impaired kidney function. With decreasing kidney function, levels of serum holoTC rise to ensure sufficient holoTC is delivered into cells. Therefore, in patients with renal insufficiency (assessed by measurement of serum creatinine), serum holoTC concentrations may appear normal, and are no longer a useful biomarker to predict B12 status (Herrmann et al., 2005). Other conditions that might elevate serum holoTC concentrations include liver diseases and possibly the development of autoantibodies against transcobalamin.

High folic acid intake has been associated with reduced serum holoTC concentrations. This adverse effect on B12 status has only been implicated with synthetic folic acid and not with natural folate from food sources (i.e., methyltetrahydrofolate) (Selhub et al., 2022). However, in countries with mandatory folic acid fortification without B12 fortification, a high-folate-low vitamin B12 interaction may be cause for concern (Sobczynska-Malefora et al., 2021).

Dietary B12 intakes affect serum holoTC concentrations with, as expected, lower levels in vegans than omnivores due to their lower B12 intake. Serum holoTC is a better biomarker of recent B12 intake than serum B12 (or methymalonic acid or tHcy) because serum holoTC concentrations increase much faster after ingestion of B12 from a meal (i.e., 6 hours) or crystalline B12 from supplements.

Interpretive criteria

Available serum holoTC assay methods, unlike serum B12, give similar values according to a European comparison study (Morkbak et al., 2005). However, to date there is insufficient data to account for the slight differences in serum holoTC concentrations by age, sex, and possibly race for adults. Furthermore, uncertainty exists about the appropriate serum holoTC cutoff to diagnose B12 deficiency. For adults, the cutoff for serum holoTC, based on the relation of holoTC to serum methylmalonic acid, is <35-40pmol/L. according to the BOND group. They provide no cutoff for deficiency for children. However, for early childhood, the BOND group set tentative normal reference ranges for four age groups from birth to 24 months of age; see Table 22b.4 (Allen et al., 2018).

Currently the BOND group have set a tentative range indicative of adequate/normal B12 status of 40-150 or 40-200pmol/L for

healthy adults, irrespective of age and the assay used.

A summary of reference intervals for serum holoTC using the most common assays are shown in Table 22b.5.

Table 22b.5 Selected reference intervals for holotranscobalamin (holoTC) with the use of the most common assays for holoTC meas

[1] Radioimmunoassay: precipitation of transcobalamin on antibodycoated beads and measurement of vitamin B12 with the use of an

[2] Microbiology: precipitation of transcobalamin by antibodycoated beads and measurement of trapped vitamin B12 with the use of a use of ELISA after removal of apotranscobalamin with vitamin B12–coated beads.

[3] ELISA, enzyme-linked immunosorbent assay.

[4] Direct: measurement of holoTC by ELISA, with the use of an antibody that is specific for holoTC.

Method	Sample (µL)	Reference range (pmol/L)
[1] Radio-immunoassay	400	24–160 37-170
[2] Microbiology	<150	42–160
[3] ELISA	100	40-150
[4] Direct	200	19–130 36-220

Based on these data, a reference interval of 40-200pmol/L is often used as indicative of the normal range for healthy adults. Nevertheless, where possible, it is preferable to establish a local reference interval for interpreting results in population-based studies and in clinical practice.

Laboratories nowadays use a diagnostic strategy for B12 deficiency that involves several biomarkers with serum holoTC or B12 as the initial test, followed by serum methylmalonic acid or homocysteine as the second line test. See Section 22b.8 on multiple biomarkers for more discussion.

Measurement of serum holo transcobalamin

Several methods are available for measuring serum holoTC. They include a radioimmunoassay (RIA) method based on monoclonal antibodies against transcobalamin (Vu et al., 1993) and an enzyme-linked immunosorbent assay procedure that is easier to use and measures holoTC directly without sample pretreatment (Brady et al., 2008). The holoTC-RIA method has now been replaced by an assay that uses holoTC-specific monoclonal antibodies and yields results comparable to those of the holoTC-RIA. An automated assay has been developed for this method using the Abbott AxSYM immunoassay analzyer (Brady et al., 2008). Tests have confirmed that results based on the holoTC-RIA and the holoTC-enzyme-linked immunosorbent assay yield similar values (Morkbak et al., 2005).

Blood sampling and storage conditions have limited effect on serum or plasma holoTC concentrations. Therefore, no special precautions are necessary for drawing blood samples and concentrations are stable in serum stored at –20°C to –70°C for at least 16 months. Both serum and plasma can be used for the assay. Fasting and non-fasting blood samples can be used because holoTC concentrations are unaffected by the intake of a normal diet (Nexo & Hoffman-Lucke, 2011).

22b.2.3 Deoxyuridine suppression test

This sensitive in vitro test has been used to diagnose early vitamin B12 (or folate) deficiency even in the absence of morphological changes in the blood. It was developed by Killman (1964) and Metz et al. (1968). Abnormal deoxyuridine suppression is the biochemical expression of disordered DNA metabolism. Although either lymphocytes or whole blood can be used for this test to detect past vitamin B12 or folate status, generally bone marrow cells are preferred because they measure the acute status (Colman 1981). The test is rarely used today because it requires bone marrow cultures, uses a radioactive label, is difficult to

control (Carmel et al., 1996; Chanarin & Metz, 1997), and is not specific to vitamin B12 deficiency.

22b.2.4 Serum methylmalonic acid

Methylmalonic acid (MMA) accumulates in the serum or plasma when the supply of vitamin B12 is reduced; it does not rise in folate deficiency (Savage et al., 1994). Hence, the measurement of methylmalonic acid in serum or plasma, is a sensitive and specific marker of tissue vitamin B12 deficiency, reflecting stores rather than intake. Concentrations are elevated in vitamin B12 deficiency, even in the absence of clinical signs or symptoms or of morphological changes in the blood. For example, in a cross-sectional study of 2919 Dutch elderly people with elevated serum homocysteine (tHcy) levels (i.e., >12 mol/L), the association of total serum B12 levels with serum MMA (and tHcy) was explored with restricted cubic splines (Figure 22b.5).

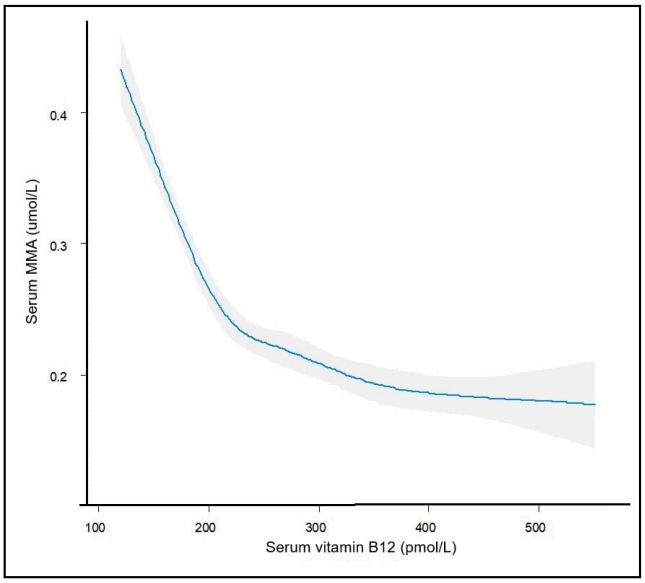


Figure 22b.5 Associations between Serum vitamin B12 and Serum MMA, adjusted for age, sex and creatinine levels. Modified from van Wijngaarden et al., 2017

Of the participants, 50% were women and all were >65 years (mean age of 74.1 years) with a mean BMI of 27.1. Even with concentrations of serum total B12 below about 330pmol/L, a rise in both serum MMA (and tHcy) levels was observed, with an even steeper increase when total serum B12 levels fell to below 220pmol/L. These findings are in line with other studies, which have observed inflections, or change points with serum total B12 concentrations between 200 and 500pmol/L. In the Dutch study, serum MMA (and serum tHcy) also rose when serum HoloTC concentrations fell to less than about 100pmol/L (van Wijngaarden et al., 2017). An additional finding was the significant correlations reported between vitamin B12 intake and total serum B12, methylmalonic acid and holoTC.

In some cases, elevated serum MMA (and tHcy) concentrations have been reduced by vitamin B12 therapy, thus confirming vitamin B12 deficiency (Lindenbaum et al., 1990; Joosten et al., 1993). For example, Bolann et al. (2000) noted that in patients (n = 196) with elevated MMA concentrations in both serum and urine (>376 nmo/L; >0.38 >mol/24 h, respectively) MMA concentrations in serum were reduced by more than 50% after vitamin B12 supplementation.

Table 22b.6 Plasma levels of cobalamin and metabolites in macrobiotic and healthy control infants. MMA, methylmalonic acid; Hcy, h Schneede et al. (1994).

	Macrobiotics (n = 41)	Controls (n = 50)	Signific
MMA (µmol/L)	1.44 (0.17-12.15)	0.18 (0.06-0.51)	$p < 0.0$
Hcy (µmol/L)	13.5 (6.8-26.8)	7.59 (5.3-11.0)	$p < 0.0$

	Macrobiotics (n = 41)	Controls (n = 50)	Signific
Cobalamin (pmol/L)	141 (59-340)	399 (194-821)	$p < 0.0$
Cysteine (µmol/L)	164 (119-226)	184 (145-233)	$p = 0.0$
Methionine (µmol/L)	6.0 (1.5-23.8)	6.9 (1.1-41.9)	$p = 0.4$

The validity of serum or plasma methylmalonic acid concentrations as a sensitive and specific test for the diagnosis of nutritional vitamin B12 deficiency during infancy has also been explored. The results of a study of 41 Dutch infants on macrobiotic diets and 50 healthy omnivorous controls (Table 22b.6) showed markedly higher plasma methylmalonic acid concentrations in the infants on macrobiotic diets compared to the controls (Schneede et al., 1994). Moreover, plasma methylmalonic acid concentrations were inversely related to plasma vitamin B12 levels. Logistic regression showed that methylmalonic acid followed by total homocysteine and cobalamin in plasma, in that order, were the strongest predictors of vitamin B12 deficiency in the macrobiotic group.

Several other investigators have compared the sensitivity of serum or plasma methylmalonic acid in diagnosing vitamin B12 (or folate deficiency) with that of serum or plasma total homocysteine (see Section 22a.2.2 for more details). Results of a study by Savage et al. (1994) (Table 22b.7) emphasize the high sensitivity of both serum methylmalonic acid concentrations and total homocysteine levels for the diagnosis of functional vitamin B12 deficiency.

Table 22b.7 Performance of serum metabolite assays in patients with clinically defined cobalamin or folate deficiency. Data from Sa

Serum level	Prevalence in cobalamin deficient patients (%)	Prevalence
Elevated MMA	98	12
Elevated Hcy	96	91
Elevated MMA, Hcy normal	4	2
Elevated Hcy, MMA normal	1	80
Hcy and MMA both normal	0.2	7

Serum methylmalonic acid concentrations are not only a sensitive biomarker but also a very specific biomarker of functional vitamin B12 deficiency. Only a few conditions confound their use as outlined below.

Factors affecting serum methymalonic acid

Age tends to increase serum MMA concentrations, particularly in persons >70y. This trend was reported in the US NHANES 2013-2014 survey of adults >20 years, with levels about 60% higher in those aged >70y compared to the youngest age group (i.e., 20-39 y) (Mineva et al., 2019). It is of interest that this age-related trend in serum MMA levels was apparent even among persons in the NHANES 2013-2014 survey with serum total B12 concentrations indicative of replete B12 status (i.e., >300pmol/L) and with normal renal function, as shown in Table 22b.8.

Table 22b.8 Serum MMA (mmol/L), sample size, medians, and central 95% reference intervals, by age group in US persons ≥20y, vi normal renal function. NHANES 2011-2014. Data from Mineva et al., 2019.

Age group (y)	Sample size	Median (95% CI)	2.5th-97.5th pe
All	5481	130 (127, 133)	67.2-281

Age group (y)	Sample size	Median (95% CI)	2.5th-97.5th pe
20-39	2178	120 (118, 124)	63.5-254
40-59	1988	133 (128, 138)	70.5-293
60-69	824	143 (134, 149)	72.4-281
≥70	491	161 (151, 167)	84.3-317

This age-related trend is said to reflect the increasingly inadequate absorption of vitamin B12 that occurs with aging (Stabler & Allen, 2004), which arises from the gradual decrease in both gastric acidity and production of intrinsic factor (Asselt et al., 1998).

Ethnicity affects serum MMA. In both the US NHANES 2011- and 2013-2014 surveys, non-Hispanic white persons had higher serum MMA (about 25%) concentrations than non-Hispanic black persons, even after adjustment for other covariates (Mineva et al., 2019).

Pregnancy affects serum/plasma MAA. In a longitudinal study of healthy women during pregnancy, after correction for hemodilution, plasma MMA concentrations showed a moderate increase throughout pregnancy from 20 weeks gestation (Murphy et al., 2007). The increase was interpreted to indicate a depletion of maternal intracellular vitamin B12 stores during pregnancy. Other reports, however, propose that an increase in plasma or serum MMA during pregnancy is normal and does not reflect a low vitamin B12 status and caution that serum MAA may not be a reliable biomarker of vitamin B12 status among pregnant women (Bae et al., 2015).

Impaired renal function, an indicator of chronic kidney disease, also leads to higher serum MMA concentrations even from the early stages of renal impairment, as shown in (Figure 22b.6).

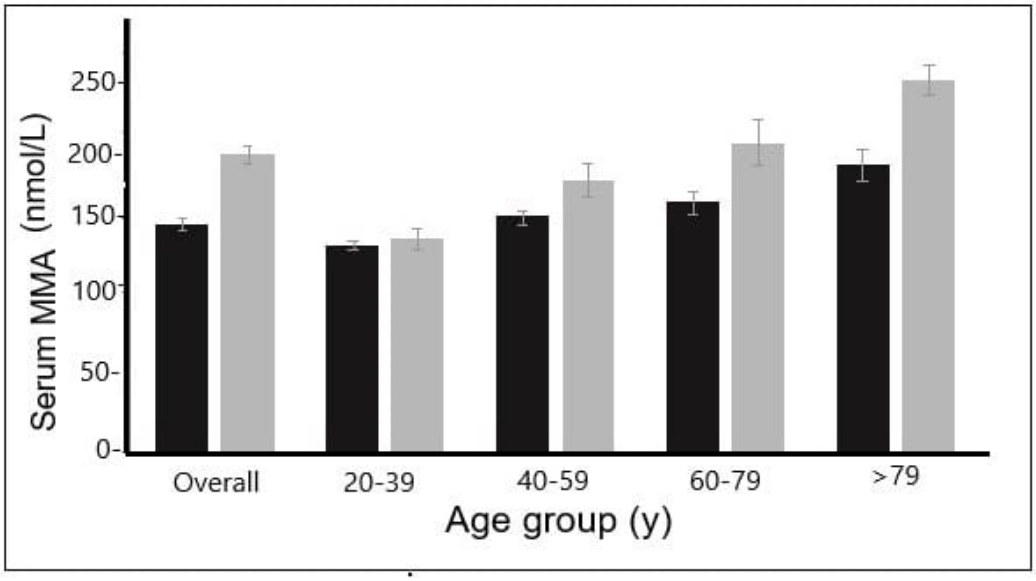


Figure 22b.6 Serum MMA geometric mean concentrations by age group and renal function in US persons ≥20y of age, NHANES 2011-2014. Black bars represent normal renal function and gray bars represent impaired renal function (stages 1-5 chronic kidney disease). Error bars represent 95% CIs. Data from Mineva et al. (2019).

Data include geometric mean concentrations of serum MMA by age group and renal function in persons ≥20y from US NHANES 2011-2014 (Mineva et al., 2019). These findings emphasize that renal function should always be assessed when measurements of serum MMA are used, especially in the elderly.

Hypovolemia (decrease in the volume of circulating blood) is associated with rising levels of serum MMA.

Hypothyroidism affects serum MMA, increasing concentrations.

Small-bowel bacterial overgrowth may elevate serum MMA concentrations due to the production of high levels of propionic acid (the precursor of MMA) by bacteria. When this condition is suspected, treatment with antibiotics should result in a decline in serum methylmalonic acid concentrations (Snow, 1999).

Inherited metabolic defects that elevate MMA concentrations in serum and urine are a group termed "isolated methylmalonic acidemia". The latter can be caused by complete or partial deficiency of the enzyme methylmalonyl-CoA mutase, which requires vitamin B12 as a co-factor (5-deoxy-adenosyl-cobalamin (Ado-Cbl)) (Figure 22b.7).

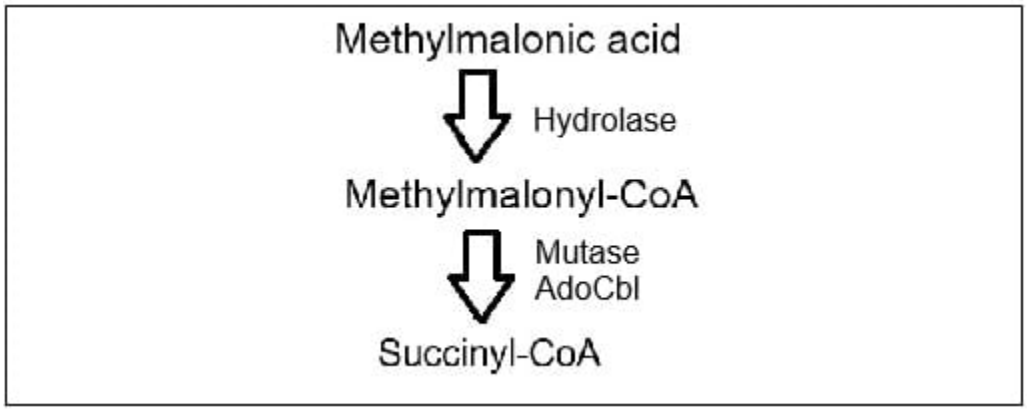


Figure 22b.7 Conversion of methylmalonylcoenzyme A (CoA) to succinyl-CoA via methylmalonyl CoA mutase (mutase).

Use of supplements of crystalline B12 reduce serum MMA concentrations. In the US NHANUES 2011-2014 survey lower serum MMA levels were reported in those participants reporting B12 supplement consumption in the past 30 days (yes or no), even after controlling for other covariates.

Fasting times may influence serum MMA concentrations. In the US NHANES 2011-2014 surveys, shorter fasting time was associated with about 10% higher serum MMA concentrations (Mineva et al., 2019).

Interpretive criteria

Currently there is no consensus for the cutoff value for serum MMA acid concentrations indicative of vitamin B12 deficiency. Two tentative “cutoffs” for deficiency in adults are presented in Table 22b.4 from the BOND group (Allen et al., 2018). These so called cutoffs are derived from serum MMA concentrations compiled from US NHANES 2003-2006 data in which the central 95% reference intervals (2.5th to 97.5th percentile) for serum MMA concentrations were presented (Pfeiffer et al., 2013).

Bailey et al. (2013) argues that use of only one serum MAA cutoff point can result in those with a B12 status considered intermediate being misclassified into the sufficient and deficient groups. To resolve this uncertainty, these investigators examined the complex relation between serum B12 and MMA (adjusted age, glomerular filtration rate, and hours of fasting) based on data from 12,683 adult participants of the NHANES 1999-2004 study. Based on statistical modeling, those participants with a serum B12 <126pmol/L had the highest serum MMA (ie., median 281nmol) whereas for those with a serum B12 >287pmol/L (who likely had adequate B12 status), serum MMA was low (i.e., 120nmol/L). They also identified an intermediate group (i.e., between those with severe B12 deficiency and those with optimal vitamin B12 status) who were defined by a serum MMA of 148nmol/L, although their vitamin B12 status was difficult to interpret.

A range of adequate/normal serum MAA concentrations for infants and children up to 2y are also provided by the BOND group

(See Table 22b.4). These values are from the same sample of well-nourished breastfed Norwegian infants used to set the age-specific adequate/normal values for serum B12. Data from the US NHANES 2011-2014 survey have been used to establish 95% reference intervals for serum MMA concentrations by age for US adults >20y. Reference intervals for serum MMA based on the overall sample are available and also on a subpopulation with serum B12 values said to be indicative of B12 repletion (i.e., ≥300pmol/L) and normal renal function based on serum creatinine (See Table 22b.8). Pregnant and lacating women were excluded in both datasets.

Not surprisingly, these age-specific central 95% reference intervals established for healthy adults (i.e., with serum B12 >300pmol/L and normal renal function) are lower than the upper-end MMA cutoffs in earlier reports whose participants were from "apparently healthy populations" but with an unknown clinical history. In these earlier reports, some older participants probably had impaired renal function that was not considered. Clearly, in studies designed to establish adequate/normal serum MAA adult values, it is advisable to evaluate renal function using serum creatinine, especially in older adults.

Note that if both serum methylmalonic acid and homocysteine are analysed, a distinction can be made between vitamin B12 and folate deficiencies. Elevated levels of both metabolites are expected in vitamin B12 deficiency, whereas in folate deficiency, only increases in serum homocysteine concentrations occur.

Measurement of serum methylmalonic acid

The analytical procedures most frequently used for serum methylmalonic acid involve liquid chromatography-tandem mass spectroscopy (LC-MS/MS). Advantages of these procedures include the small sample size needed, their high throughput, and good precision. A summary table of the characteristics of the differing LC-MS/MS procedures available is presented in Jin et al. (2022). In some cases, the LC-MS/MS procedure has been used to analyze MMA in serum, urine, or dried blood spots. The major drawshy;back of the LC-MS/MS procedures is their cost, the availability and maintenance of the equipment, and expertise required (Allen et al., 2018).

A more cost-effective method requiring small volumes of reagents with high throughput has been developed for serum MAA using stable-isotope-dilution LC-MS/MS. Application of this method will facilitate early identification of vitamin B12 deficiency among the elderly (Jin et al., 2022). Levels of MMA in serum are stable for several years, when samples are stored at −70°C.

22b.2.5 Urinary methylmalonic acid excretion

Vitamin B12 serves as a cofactor for methylmalonyl CoA mutase (EC 5.4.99.2) as noted earlier.This enzyme acts in the conversion of methylmalonyl CoA to succinyl CoA (Figure 22b.7). Therefore, in vitamin B12 deficiency, methylmalonic acid accumulates in the blood and is excreted by the kidneys in increased amounts in the urine, resulting in methylmalonic aciduria (Norman & Morrison, 1993).

The biomarker methylmalonic acid in urine is sensitive, detecting early functional vitamin B12 deficiency before the onset of macrocytic anemia, as noted for MMA in serum. Similarly, this test is also highly specific and is not affected by folate deficiency. Only rare congenital enzyme defects discussed earlier (i.e., methylmalonic aciduria) interfere with the test.

In elderly patients with renal disease, methylmalonic acid accumulates so levels in urine, like serum, are elevated leading to a decreased specificity for B12 deficiency. In such cases, MMA concentrations in urine should be adjusted for creatinine excretion (Norman & Morrison, 1993) and the ratio of urinary MMA to creatinine (uMMA/C) applied to assess B12 status. Such an approach is recommended in view of the finding that uMMA/C is not dependent on renal function (Supakul et al., 2020). Normal levels for serum creatinine range from 0.6-1.1mg/dL for women and 0.7-1.3mg/dL for men(Allen et al., 2018).

Interpretive criteria and measurement of methylmalonic acid in urine

The reduced specificity of plasma MMA during renal impairment has led to the development of the uMMA/C ratio in the diagnosis of B12 deficiency. A threshold for uMMA/C ratios of 1.45µmol/mmol has been established for adult patients with plasma B12

concentrations in the low-normal range (i.e., 201-350ng/L). This threshold has good diagnostic performance for B12 deficiency in patients with subnormal serum or plasma B12 concentrations and renal impairment (Supakul et al., 2020).

Assay of urine for MMA avoids the need for blood collection, although excretion of urinary MMA increases after meals. Analysis of 24h urine samples avoids this problem, but collection of 24hr urine samples is hard to achieve. Instead, analyses of two casual urine samples to correct for variability in the MMA concentration in urine can be used.

Combined gas/ liquid - chromщatography mass-spectrometry techniques (LC-MS/MS or GC-MS/MS) are used to measure MMA in urine, as discussed for serum MMA. They are sensitive and reliable (Norman & Morrison, 1993; Allen et al., 2018) but the procedure is technically difficult; the equipment is expensive, needs high maintenance and expertise. The test is best performed on a 24h urine sample.

Recently a method has been developed and validated in which MMA/creatinine can be assayed in fasting urine samples collected on filter paper. After eluting the dry urine samples from the filter paper with a solution containing internal standards followed by filtration, the MMA/creatinine ratios are measured by ultra-performance LC-MS/MS. The dried urine samples are stable at 86 days at room temperature and can be used for screening B12 deficiency in the elderly (Boutin et al., 2022).

Methylmalonic acid in urine is stable at room temperature, provided concentrated hydrochloric acid is added to the urine sample, and for several years, when urine specimens are frozen at –70°C.

22b.2.6 Red blood cell methylmalonic acid

Unrecognized deficiency of vitamin B12 during early childhood can lead to long-term health outcomes in children including impaired cognitive function and developmental delay. Maternal B12 deficiency during pregnancy is the most common cause of vitamin B12 deficiency during early infancy, as both B12 fetal stores and vitamin B12 concentrations in breast milk are low. During later infancy and early childhood inadequate intakes of B12 due to low intake of animal-source foods and possibly malabsorption, may play a role (Sobczynska-Malefora et al., 2023).

Methylmalonic acid is considered the most specific functional biomarker and a sensitive indicator for B12 deficiency, as noted earlier. In many countries, analysis of dried blood spots are used as a diagnostic tool for newborn screening. In many countries, however, even though dried blood spots are collected routinely from newborns, the blood spots are not routinely screened for B12 status. Therefore, developing a sensitive assay of MMA in dried blood spots through a newborn screening programme has the potential to facilitate the timely diagnosis and treatment of neonatal B12 deficiency. Such a method in dried blood spots could also facilitate testing B12 status in large-scale surveys and in remote populations (Schroder et al., 2014). Details of the assay of MMA in dried bloods spots and their interpretation are given below.

Interpretive criteria and measurement of methylmalonic acid in dried blood spots

A new highly sensitive procedure based on stable isotope dilution with liquid chromatography-tandem mass spectrometry (LC-MS/MS) has been used to establish a reference interval for MMA in dried blood spots for healthy term newborn infants. Details of the method are available in Schroder et al.(2014). The mean dried blood spot MMA concentration quantified from 160 newborns was 16.8pmol/L (95th CI:15.9-17.6pmol/L per 8-mmspot), with 95% reference intervals (2.5th to 97.5th percentile) for MMA of 9.89 to 29.3pmol/8-mm blood spot (0.450 to 1.33µmol/L whole blood). Concentrations of MMA above the upper limit of this reference interval indicate an increased risk of B12 deficiency. The establishment of this neonatal MMA reference interval in dried blood spots will assist in the timely diagnosis and treatment of neonatal B12 deficiency through newborn screening programmes. To date, comparable reference intervals for MMA in dried blood spots of adults have not been established.

Concentrations of MMA in dried blood spots increase with storage at room temperature so care must be taken to store dried blood spots below –80°C prior to analysis. Nevertheless, because storage at room temperature for at least one week has no effect on the dried blood spot MMA concentrations, they can be shipped at room temperature (Schroder et al., 2016).

22b.2.7 Measurement of vitamin B12 absorption

Once vitamin B12 deficiency has been diagnosed, it is often of interest to establish if malabsorption is the cause. Originally, the Schilling test was used for this purpose, but it is no longer used in view of the safety factors related to the use of radioactive B12 (^{57}Co-cyanocobalamin) (Sobczynska-Malefora et al., 2021). Instead, alternative tests are now employed that measure either the change in serum holoTC (CobaSorb) concentrations or the absorption of microbially produced cobalamin labeled with ^{14}Carbon or ^{13}Carbon. However, the tests employing both ^{14}Carbon or ^{13}Carbon labeled cobalamin require more validation to diagnose B12 malabsorption in clinical patients (Miller & Green, 2020).

CobaSorb test depends on the fact that newly absorbed B12 circulates in the plasma as holoTC. The test is used to clarify whether the B12 deficiency diagnosed in a person is caused by an inability to absorb the vitamin. It is important that the test is performed before participants are treated for B12 deficiency in order to detect an increase in serum holoTC. The test is based on the analysis of serum holoTC prior to and following an oral dose of unlabeled B12. For the test, first a blood sample is taken to measure plasma holoTC, after which a dose of 9µg of crystalline B12 in water is given orally at 6h intervals over a 24h period. The next day, a second blood sample is taken, so that the change in the amount of B12 bound to serum holoTC can be determined. A small or no increase in holoTC is indicative of malabsorption. The test measures relative absorption and does not provide a quantitative estimate of B12 bioavailability Brito et al., 2018). Both the limited use of holoTC assays and lack of interpretative knowledge in many clinical settings have prevented widespread use of the CobaSorb test.

^{14}C-labeled B12 absorption test uses carbon-14-labeled vitamin B12 (^{14}C-B12) produced by growing a genetically modified strain of *Salmonella enterica* in a medium containing ^{14}C dimethybenzimidazole. After oral ingestion of ^{14}C-B12, enrichment of ^{14}C in blood, urine, and stool samples can be measured using accelerator mass spectrometry. This approach has been used to assess absorption and bioavailability of B12 from aqueous solutions as well as some endogenously enriched foods. Exposure to radioactivity is negligible because the enrichment levels of ^{14}C-B12 can be measured in very small amounts (microliters) of the biological samples (Carkeet et al., 2006).

^{13}C-labeled B12 absorption test is a nonradioactive method for determining vitamin B12 absorption and bioavailability. ^{13}C-cyanocobalamin is a safe stable isotope-labeled vitamin B12 also biosynthesized from *Salmonella enterica* (Devi et al., 2020). For this test, a baseline blood sample followed by serial blood samples between 5 and 7h post oral administration of the ^{13}C-cyanocobalamin are required. After processing, enrichment of plasma concentrations of ^{13}C-cyanocobalamin is assessed at 5 and 7h post dose by ultra-HPLC-MS. The kinetics of the plasma appearance of ^{13}C-cyanocobalamin in a 2-compartment model provide an index of oral absorption and bioavailability of B12. Both the ^{14}C and ^{13}C labeled absorption tests show promise but have not yet been developed for clinical tests, requiring further testing and validation (Brito et al., 2018).

Alternative procedures that can assist in the diagnosis of malabsorption from classic pernicious anemia arising from an autoimmune disease or from long-term atrophy or atrophic gastritis, include a combination of anti-intrinsic factor antibodies, and gastrin and pepsinogen A and C in serum (Green et al., 2017). These tests can all be performed using serologic assays, and if used in combination they can identify most cases of malabsorption from pernicious anemia (Lindgren et al., 1998).

22b.2.8 Multiple biomarkers

Defects in the hematopoietic system do not generally occur until the final stage of B12 deficiency. At this stage, the hematological defects are indistinguishable from those of folate deficiency and include megaloblastic anemia, characterized by macro-ovalocytic erythrocytes, abnormal red cell indices, and a low hemoglobin concentration (see Chapter 17, Sections 17.2.7 and 17.4). However, identifying impaired vitamin B12 status in the earlier stages is more challenging, but can be enhanced by using one or more plasma biomarkers of B12 status. No single biochemical biomarker is suitable for diagnosing B12 deficiency in all persons, and the use of several biomarkers is preferred. For a useful summary table comparing the relative strengths and weaknesses of the B12 biomarkers, see Allen et al. (2018). Algorithms based either on a selection of biomarkers measured sequentially or, multiple biomarkers analyzed simultaneously and then combined into a single diagnostic indicator, are both approaches that are used to circumvent the inherent limitations of each individual biomarker.

Algorithms based on the selection of sequential assays are used by several laboratories to assess B12 status. For example, initially, serum total B12 or serum holoTC can be measured, and if the concentration is deemed low (i.e., serum B12<148pmol/L or serum holoTC<35pmol/L), then B12 deficiency is diagnosed.

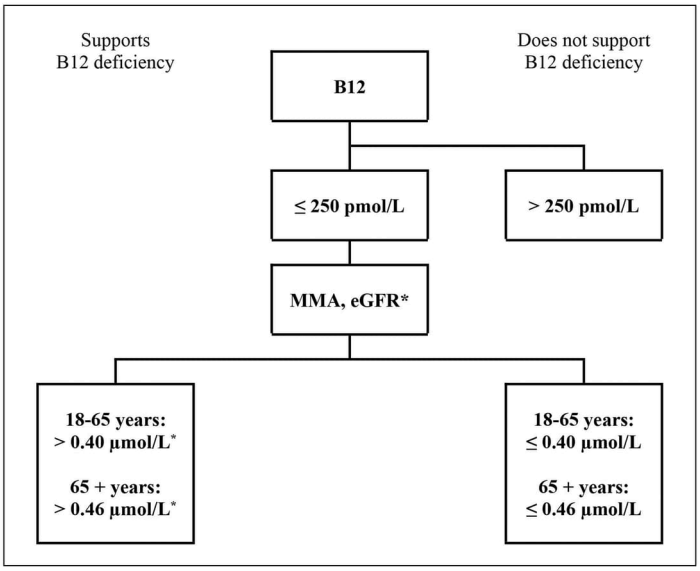


Figure 22b.8 Biomarkers indicating B12 deficiency in Denmark (Nexo & Parkner, 2024). Total serum B12 and MMA are included in the figure. HoloTC may replace total B12 as a first line test, and homocysteine may be used instead of MMA, notably in folate replete populations. * For patients with an eGFR below the reference interval (60mL/min), an elevated serum MMA concentration may be caused by the impaired kidney function rather than B12 deficiency. B12 represents total plasma B12; MMA, methylmalonic acid; eGFR, estimated glomerular filtration rate.

However, if the concentration of either serum B12 or holoB12 is in the intermediate range between inadequate and adequate B12 status (i.e., between 148 and 250pmol/L for serum total B12 or between 35 and 50pmol/L for serum holoTC), then a functional biomarker such as serum methymalonic acid or serum homocysteine is measured as the second-line test. If either serum methylmalonic acid or homocysteine is considered elevated, then B12 deficiency is diagnosed.

As an example, algorithms employed in Denmark are presented in (Figure 22b.8). Slightly different procedures are used in The Netherlands. For both algorithms, cutoffs indicative of B12 deficiency are selected, generally from the lower limits of the local reference intervals.

Table 22b.9 presents commonly seen vitamin B12 biomarker patterns in selected clinical scenarios (Sobczynska-Malefora et al., 2021).

Table 22b.9 The commonly seen vitamin B12 marker patterns in selected clinical scenarios. Cbl: cobalamin; holoTC: holotranscobalacysteine. Modified from Sobczynska-Malefora et al., 2021.

Serum holoTC	Serum total B12	Plasma MMA	Plasma tHcy
within ref. range	within ref. range	elevated	elevated
decreased	decreased	within ref. range or elevated	elevated
within ref. range	within ref. range	elevated	within ref. range
very very low	very very low	highly elevated	highly elevated
within ref. range or decreased	decreased	within ref. range or elevated	within ref. range

Serum holoTC	Serum total B12	Plasma MMA	Plasma tHcy
decreased	very low	elevated	within ref. range or elevated
within ref. range	within ref. range	within ref. range	elevated to very highly elevated
within ref. range or elevated	very low	within ref. range	within ref. range
very very low	within ref. range or decreased or elevated	highly elevated	highly elevated
within ref. range	within ref. range	very highly elevated	very highly elevated
within ref. range	within ref. range	very highly elevated	within ref. range
within ref. range	within ref. range	within ref. range	very highly elevated
within ref. range or decreased	within ref. range or decreased	within ref. range or elevated	very highly elevated
highly elevated	highly elevated	within ref. range	within ref. range
very highly elevated	very highly elevated	within ref. range	within ref. range

Concurrent measurement of several biomarkers is an alternative to sequential approaches. In the original mathematical model four biomarkers are combined: serum holoTC, serum B12, serum methylmalonic acid (MMA), serum homocysteine (Hcy) to yield a combined indicator termed the cB12 or 4cB12. The combined four-component algorithm adjusts for age and folate status, in view of the likely increase in serum MAA with age in the elderly, and the increase in serum Hcy in folate deficiency. 4cB12 can be calculated:

4cB12=log10[(holoTC.B12)/MMA.Hcy)]-(age factor)

The model 4cB12 is currently considered the best biomarker stratification for B12 status and yields one of five diagnoses: elevated B12, adequate B12, decreased B12, possible B12 deficient, and probably B12 deficient. The advantage of 4cB12 is that it is independent of local reference values and can be adjusted to correct for folate status and age. However, cost is a major disadvantage of the 4cB12 model as well as lack of availability of all four tests in routine clinical practice.

Suggested cutoffs and the interpretation of values of the combined 4cB12 indicator for use by researchers and clinicians and for epidemiologic purposes are shown in Table 22b.10.

Table 22b.10 Suggested cutoffs and interpretation of values of the combined indicator 4cB12 From Fedosov et al. (2015). B12, vitamin B12; Hcy, homocysteine; holoTC, holotranscobalamin; MMA, methylmalonic acid.

For researchers and clinicians			For epidemiologic purposes	
Classification	Biological interpretation	Guidelines	Class	Guidelines
Elevated B12 >1.5	The pathogenesis of high B12 is not fully understood	Consider possible causes of high concentrations such as liver disease or current or recent supplementation or treatment	B12 adequacy > −0.5	No action
B12 adequacy −0.5 to 1.5	Expected to accomplish all B12 status–dependent functions	No action advised		
Low B12 −1.5 to −0.5	Potential subclinical manifestations of B12 deficiency, i.e., absence of hematologic changes, but subclinical neurologic impairment	Consider recommending oral supplements	Transitional B12 status −0.5 to −2.5	Need fortification

	For researchers and clinicians		For epidemiologic purposes	
Possible B12 deficiency < −2.5	Potential manifestations of B12 deficiency	Potentially prescribe oral supplements, assess again in 3-6mo	Low B12 status < −2.5	Need supplementation, rechecking, monitoring of status over time
Probable B12 deficiency < −2.5	It is possible to observe clinical manifestations of B12 deficiency. Clinical outcomes are needed to confirm potential clinical deficiency	Consider immediate treatment with i.m.injections, determine cause, consider particularly possible pernicious anemia		

So far these cutoffs for 4cB12 have been validated against hemoglobin concentrations and cognitive status (based on the Mini Mental State Examination), although additional validation studies using other cognitive and neurological outcomes are needed. As yet, the 4cB12 model has not been validated for use in pregnancy but has proven very useful in the research setting (Fedosov et al., 2015).

More recently, in view of the drawbacks of 4cB12, refined models have been developed based on "missing" biomarkers. These alternative cB12 equations, missing one (3cB12) or two (2cB12) biomarkers of B12 status, have been compared with the established four-parameter 4cB12. Revised cut-points and guidelines for using this approach can be found in Fedosov et al. (2015) and Campos et al. (2020). The missing biomarker methods need further validation, especially for use in pregnancy and early childhood.

This page titled 22.2: Biomarkers of vitamin B12 status (22b.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23: CALCIUM (CHAPTER 23A)

Abstract

Calcium is known for being a major constituent of bones and teeth. About 99% of the body's calcium is found in these hard tissues. Calcium is also an extracellular divalent cation in the blood, where it exerts functional roles such as cell signaling and membrane transport. Its concentration in blood is tightly regulated by many hormones including parathyroid hormone (PTH), calcitriol, and calcitonin. Blood levels of calcium are not useful in determining the nutritional status of calcium because of these hormonal influences. Dietary requirements for calcium are derived using calcium balance, which is negative when calcium intake is too low to maintain normal bone density and zero in adults when calcium intake is sufficient or positive during growth and pregnancy. Calcium balance can be estimated using dual-absorptiometry (DXA) or measuring urine and fecal losses. A calcium intake above requirement does not stimulate further increases in bone mass; instead excess intake can lead to adverse effects such as soft tissue calcification and kidney stone formation. When calcium intakes are low, and/or when vitamin D is inadequate to permit sufficient calcium absorption, bone health is impaired leading to rickets in children and osteomalacia or osteoporosis in adults. Other health benefits of calcium are under investigation.

23.1: Calcium (23a.1)

23.2: Functions of calcium (23a.2)

23.3: Absorption and metabolism of calcium (23a.3)

23.4: Dietary sources and intakes of calcium (23a.4)

23.5: Nutrient reference values for calcium (23a.5)

23.6: Calcium deficiency (23a.6)

23.7: Effects of high intakes of calcium (23a.7)

23.8: Biochemical indices of calcium status (23a.8)

23.9: Measurement of bone mineral content and bone mineral density (23a.9)

This page titled 23: Calcium (Chapter 23a) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.1: Calcium (23a.1)

Calcium is the most abundant mineral in the body, accounting for 10–20g per kg body weight in adults. Bone is the main reservoir for calcium. The remainder is located in body tissues (10g), blood, and the extracellular fluids (900mg). In contrast, a greater proportion of the distribution of phosphorus and magnesium, the other two bone minerals, is located in soft tissue (Table 23a.1). All of these three minerals have two major roles — as structural components in bone and regulatory agents in body fluids.

Table 23a.1: Calcium, magnesium and phosphorus distribution in an adult weighing 70 kg

Element	Mass (g)	Percentage in the skeleton	Percentage
Calcium (Ca)	1200	99	1
Magnesium (Mg)	24	55	40
Phosphorus (P)	900	80	12

This page titled 23.1: Calcium (23a.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.2: Functions of calcium (23a.2)

Calcium has a structural role in bones and teeth, where 99% of the body's calcium exists primarily as calcium hydroxyapatite [$Ca_{10}(PO_4)_6(OH)_2$]. Bone is a dynamic tissue that undergoes remodeling throughout life except for the teeth. Bone formation is under the control of osteoblasts, the bone forming cells, whereas bone resorption is initiated by osteoclasts, the bone resorbing cells. During growth, bone formation exceeds resorption as bone modeling is predominant. In later life when only bone remodeling takes place, bone resorption begins to occur at a faster rate than bone formation and this results in age-related bone loss (Siddiqui and Partridget, 2016).

The remaining 1% of total body calcium is in the soft tissues where it has a regulatory role in several metabolic processes, including enzyme activation, vascular contraction and vasodilation, muscle contractibility, nerve transmission, hormone function, and membrane transport. When the levels of absorbed calcium from the diet are insufficient to balance the obligatory fecal and urinary losses, calcium is drawn from the bone and blood calcium levels are maintained within narrow limits. This homeostatic regulation of the calcium level in the blood is achieved by the interaction of the parathyroid hormone (PTH), 1,25-dihydroxyvitamin D (1,25$(OH)_2$D), and calcitonin, the three major calcium-regulating hormones (Peacock, 2010). Phosphorus, protein, estrogen, and other hormones, including glucocorticoids, thyroid hormones, growth hormone, and insulin also have a role (Siddiqui and Partridget, 2016).

This page titled 23.2: Functions of calcium (23a.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.3: Absorption and metabolism of calcium (23a.3)

Calcium balance refers to the state of the calcium body stores (in bone) which are a function of dietary intake, intestinal absorption, renal excretion, and bone remodeling. A major control of calcium balance is the amount of calcium absorbed. Calcium is absorbed across the intestinal mucosa in two ways: active transport and passive para-cellular diffusion. Active absorption of calcium is influenced by calcium and vitamin D status of the individual, age, pregnancy, and lactation. Active absorption occurs when intakes of calcium are low relative to need, and is regulated primarily via $1,25(OH)_2D$ and its intestinal receptors. Passive absorption is more important at high intakes of calcium (Weaver and Peacock, 2019).

Efficiency of calcium absorption is highest when dietary calcium intakes are low or during periods of rapid growth in infancy, early childhood, and adolescence, and during pregnancy and lactation. Indeed, absorption is very high during infancy (about 60%) (Abrams et al., 1997) and about 34% during puberty (Abrams and Stuff, 1994), stabilizing at about 25–30% in young adults. Subsequently, the intestinal absorption of calcium decreases progressively with age (Christakos et al., 2011).

Urinary calcium excretion is also regulated in response to need. Normally calcium re-absorption in the kidney nephrons is not 100%. When there is hypocalcemia to stimulate PTH release, one action of PTH is to increase efficiency of calcium re-absorption thus reducing calcium losses in urine(Peacock, 2010). Several dietary variables affect both calcium absorption and retention and, hence, calcium status. Diets high in protein increase the urinary excretion of calcium (hypercalciuria), which, however, is compensated by increased calcium absorption (Mangano et al., 2014). Dietary sodium also influences urinary calcium losses, with high sodium intakes increasing urinary calcium losses (Bedford and Barr, 2011), and lowering bone density in women (Devine et al., 1995). Caffeine, like protein and sodium, also increases urinary calcium loss but has only a small negative effect on calcium retention, which can be offset by increasing dietary calcium intakes (Barrett-Connor et al., 1994).

High phosphorus intakes reduce urinary calcium losses (i.e., have a hypocalciuric effect), but because they increase losses of endogenous fecal calcium at the same time (Heaney and Recker, 1994), their net effect on calcium balance is still under debate. One might assume that there would be no effect except in conditions in which the coordination between calcium and phosphate homeostasis occurs through the actions of PTH, FGF-23 and $1,25(OH)_2D$. The disruption of this coordination, however, can occur with disease states such as chronic kidney disease (CKD) (Peacock, 2010).However, increasing dietary phosphorus through inorganic phosphate additives may have detrimental effects on bone and mineral metabolism in humans and animals (Vorland et al., 2017).

Phytates and oxalates can inhibit the absorption of calcium by forming insoluble calcium complexes in the gastrointestinal tract (Barger-Lux et al., 1995). Hence, they may have a negative effect on calcium status, although their overall effect may be small unless dietary calcium intakes are low. Some other dietary components, including lactose and inulin, enhance passive calcium absorption by increasing calcium solubility in the ileum (Weaver et al., 2016).

Several clinical conditions are associated with disturbances in calcium absorption. Calcium malabsorption is associated with gastrointestinal diseases such as Crohn's and celiac disease, and intestinal resection or bypass. Intestinal absorption of calcium is also impaired in patients with renal failure, caused by the reduced synthesis of 1,25-dihydroxyvitamin D (Peacock, 2010). Excessive intestinal absorption of calcium and hypercalcemia occur in sarcoidosis, when excessive calcitriol appears in serum, and also in vitamin D intoxication.

This page titled 23.3: Absorption and metabolism of calcium (23a.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.4: Dietary sources and intakes of calcium (23a.4)

Calcium is not widely distributed in all food groups, as shown in Table 23a.2.

Table 23a.2 Calcium bioavailability chart showing calcium content, estimated absorption, and net amount absorbed.
** Theoretical amount of calcium absorbed by an adult. Source: Weaver CM, Plawecki KL (1994). Dietary calcium: adequacy of a vegetarian diet. American Journal of Clinical Nutrition. 59: 1238S-1241S.*

Food	Serving size	Calcium Content (mg)	Estimated Absorption %	Net Ca* Absorbed (mg)
Cow milk (fluid)	250mL	310	32	100
White beans (cooked)	125mL	85	22	18
Red beans (cooked)	125mL	26	24	6
Whole wheat bread	35g slice	26	82	21
Broccoli (cooked)	125mL	33	20	20
Spinach (cooked)	125mL	129	5	7
Almonds (raw, roasted)	60mL	97	21	21
Tofu (set with $CaSO_4$)	85g	171	31	53
Soy “milk” (unfortified)	125mL	5	31	2

Milk and milk products (e.g., cheeses, yoghurt) are excellent sources of readily available (bioavailable) calcium. Plant-based beverages made from soy, almonds, rice and oats are available as substitutes for milk, but must be fortified. The availability of the calcium from these beverages is unknown, except for fortified soymilk for which the calcium availability is similar to cow milk, but only if calcium carbonate is used as a fortificant (Weaver and Peacock, 2019). Unfortified plant-based beverages have little calcium. Canned fish with bones (e.g., salmon, sardines) are good sources of calcium. Leafy-green vegetables are also good sources, but absorption of calcium from some of these foods, especially spinach, may be low (Weaver and Plawecki, 1994). Ancient grains (e.g., teff, quinoa, amaranth) and nuts are good sources of calcium (Cormick and Belizan, 2019), but highly refined or modern cereal grains such as wheat and maize are not.

Preparation of foods may also be a factor influencing their calcium content. For example, in Guatemala, when corn (maize) tortillas are prepared according to the Mayan tradition of cooking corn with limestone and leaving it to soak overnight in hot water, the calcium content rises from about 10mg to about 200mg of calcium per 100g (Belizan and Villar, 1980). In Sub-Saharan Africa, dietary calcium intake may be increased by using powdered eggshell added to foods (Bartter et al., 2018).

Foods fortified with calcium are available in many countries. In the European Union, the following forms of calcium are authorized for addition to foods and for use in food supplements: carbonate, chloride, citrate malate, gluconate, glycerophosphate, lactate, hydroxide, oxide, acetate, L-ascorbate, bisglycinate, salts of citric acid, pyruvate, salts of orthophosphoric acid, succinate, L-lysinate, malate, L-pidolate, L-threonate, sulphate (EFSA, 2012). Most of these forms are allowable in other countries. The citrate malate calcium salt is an allowable fortificant for orange juice in the United States (ODS, 2020). The United Kingdom is the only country where calcium fortification is a mandatory addition to flour; 235–390mg of calcium carbonate are added per 100g of flour. This represents about 94–156mg of elemental calcium per 100g of flour (Cormick et al., 2020). Other countries permit calcium addition to certain foods, often proprietary packaged foods, a practice called discretionary fortification.

Supplemental sources of calcium are widely available in many dosages and in a variety of chemical forms in many countries. These pills, capsules and liquids are usually regulated by the government. For example, in the United States, the following forms of calcium are allowable as supplements: carbonate, phosphate, citrate, gluconate, and lactate (Weaver and Peacock, 2019; ODS, 2020). Some supplements that are derived from oyster shell, algae, and in chewable drugs labeled as antacid such as TUMS™, are mainly in the form of calcium carbonate, the form of calcium naturally occurring in rocks such as chalk and limestone. This form of calcium is generally the cheapest to buy and has the highest amount of calcium by weight (40%). However, calcium

carbonate is poorly absorbed unless stomach acid is available for dissociation. Also, calcium carbonate is the form of calcium that causes the most side-effects such as bloating and constipation (ODS, 2020).

Dietary calcium intakes vary across countries, ranging from 175–1233mg/d (Balk et al., 2017). Countries in Asia have average dietary calcium intakes less than 500mg/d while in countries in Africa and South America, calcium intakes vary from 400–700mg/d. Only Northern European countries have a national calcium intake greater than 1000mg/d, reflecting a high dairy consumption. The average calcium intake is generally lower in women than men, although across the 74 countries included in an analysis by the International Osteoporosis Foundation (IOF, 2020), no clear patterns regarding differences by age, sex, or socioeconomic status emerged.

This page titled 23.4: Dietary sources and intakes of calcium (23a.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.5: Nutrient reference values for calcium (23a.5)

New EAR, RDA and UL levels for calcium were published by the Institute of Medicine in 2011, as shown in Table 23a.3.

*Table 23a.3 Estimated Average Requirement (EAR), Recommended Dietary Allowance (RDA, or Adequate Intake equivalent) and Upper level (UL) for calcium. * Applies to pregnant and lactating women. Source: Institute of Medicine (2011) and WHO (2018).*

Dietary Reference Intakes Ca (mg/d) IOM (2011)			
Age/sex groups	**EAR**	**RDA**	**UL**
0–6 mo	–	200 (AI)	1000
6–12 mo	–	260 (AI)	1500
1–3y F & M	500	700	2500
4–8y F & M	800	1000	2500
9–18y F* & M	1100	1300	3000
19–50y F* & M	800	1000	2500
51–70y F	1000	1200	2000
51–70y M	800	1000	2000
71+y M & F	1000	1200	2000
World Health Organization (2018)			
Pregnancy		1500	

Calcium balance studies provided the evidence for values for ages 1–50y, whereas for those over 50y, values were set to prevent bone loss. For infants age 0–6mos, an AI is given based on the average intake of calcium from human milk, whereas at age 7–12mos, the AI assumes 120mg calcium from human milk plus 140mg calcium from complementary food. The values set for the AIs for infants provide enough absorbed calcium to meet the calcium requirements estimated by the factorial method. For children, estimates of desirable bone accrual were considered; EAR values, which start at 500mg for toddlers (age 1–3y), rise to 800mg for children, then peak at 1100mg for adolescents, were based on a new calculation of calcium retention data (Vatanparast et al., 2010). For adults, maintenance of bone through attaining neutral calcium balance is the goal. The EAR values are 800mg for men age 19–70y and for women 19–50y, and the RDA for these age groups is 1000mg. The difference in age cut-offs for men (i.e., 70y) and women (i.e., 50y) reflects data showing that women at 50y experience abrupt bone loss at menopause in contrast to men at this age where bone loss is slow. For women over 50y and men over 70y, EAR values are similar at 1000mg which translates into an RDA value of 1200mg. This higher level set for women over 50y and men over 70y was considered a justified public health measure in the face of supportive yet inconclusive data suggesting a higher intake of calcium would prevent bone loss and protect against fractures. Values during pregnancy and lactation are not different from other women in the same age group in view of the well documented ability of pregnant and lactating women to increase calcium absorption, and hence compensate for their increased needs (Kovacs, 2016).

The IOM Upper Level (UL) for calcium for infants 0–1y was derived from feeding data that indicated no adverse effects on calcium excretion when calcium intakes as high as 1000–1500mg were provided. For all other age groups, the UL for calcium (Table 23a.3) was based on data related to the incidence of kidney stones, largely from the research conducted with post-menopausal women in trials where calcium supplements were provided. In contrast, the European Food Safety Authority (EFSA, 2012)retained their calcium UL of 2500mg for adults, including pregnant and lactating women, on the basis that the new data did not alter their previous adult UL estimates. Further, EFSA was not convinced there were sufficient data to set a UL for infants, children, or adolescents.

This page titled 23.5: Nutrient reference values for calcium (23a.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.6: Calcium deficiency (23a.6)

23a.6.1 Hormonal responses to calcium deficiency

Calcium homeostasis is under systemic (i.e., hormonal) control together with additional local regulating factors through intestinal absorption, influx and efflux from bone, and calcium excretion and re-absorption by the kidney. Serum calcium concentrations must be kept constant at approximately 2.5mmol/L, a concentration that is controlled by the interrelated action of three hormones: parathyroid hormone (PTH), 1,25$(OH)_2$D, and calcitonin. The first two hormones determine how much calcium is retained in the body through control of absorption, bone turnover, and renal excretion. Calcitonin has a role in how calcium moves between the extracellular fluid and bone (EFSA, 2012).

Fasting, which represents acute calcium deficiency, causes a decrease in serum calcium concentrations, inducing the release of PTH via the calcium-sensing receptor (CaSR) located on the cell surface of the parathyroid glands (EFSA, 2012). In hypocalcemia, PTH has three actions which together bring serum calcium levels back to normal. The immediate response is bone resorption to release calcium into the blood. The second response is to increase renal re-absorption of calcium in kidney tubules in order to retain calcium. Finally, PTH stimulates 1,25$(OH)_2$D synthesis in the kidney (Figure 23a.1)

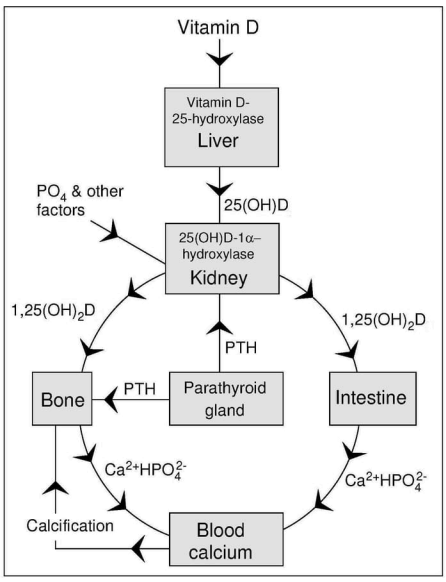


Figure 23a.1 Formation of vitamin D metabolites. Once the transport form 25(OH)D is made, the Endocrine pathway (Section 18b.2.4) for synthesis of the active metabolite 1,25$(OH)_2$D is illustrated. Parathyroid hormone (PTH) directs synthesis of 1,25$(OH)_2$D in response to a need for calcium or phosphate. Plasma 1,25$(OH)_2$D stimulates intestinal calcium transport and bone calcium mobilization, then blunts PTH synthesis to turn cycle off (Section 18b.13). Adapted from Holick, Kidney International 32: 912–929, 1987.

which serves to allow active absorption of calcium in enterocytes. Feedback control of this rise in serum calcium occurs as a normal serum concentration of calcium inhibits PTH secretion. In addition, higher than normal serum calcium stimulates calcitonin secretion by the para-follicular C cells of the thyroid gland, causing calcium to move into bone. Calcitonin's actual role in bone formation remains unclear, but it is a recognized treatment for hypercalcemia (Felsenfeld and Levine, 2015).

23a.6.2 Bone health

Chronic calcium deficiency, arising from an habitually inadequate intake or poor absorption of calcium, is one of the many important factors associated with reduced bone mass. Bone mass in later life depends on the peak bone mass achieved during growth and the rate of the subsequent age-related bone loss. The age span over which peak bone mass is achieved is between

the early teenage years and age 30, during which time 95% of bone mass is achieved. By 30y further gains are unlikely (Weaver et al., 2016), although exact timing varies with the skeletal site and sex. Optimization of bone mass is important because it reduces the risk of osteoporosis in later life.

Calcium intake and its association with bone health has been the subject of intense debate as epidemiological studies have not always found a consistent relationship between calcium intake and bone density. One reason is that calcium is a threshold nutrient so that when present in sufficient amounts, a higher calcium intake does not result in any further improvement in bone mineral accrual (Heaney, 2007). In most countries the Nutrient Reference Values (NRVs) have been set for maintenance of bone health. For example, in the United States and Canada, the Estimated Average Requirement (EAR) and Recommended Dietary Allowance (RDA) for calcium for all ages is based on maintaining a positive calcium balance in children and preventing a negative calcium balance in adults (IOM, 2011); more details are described below.

Calcium-deficiency rickets and osteomalacia occur in many parts of the world. Case reports from the United States, South Africa, and Nigeria have described infants and children with radiological rickets, growth retardation, and biochemical signs of hyperparathyroidism but with normal vitamin D status. In all these cases of rickets, habitual calcium intakes were low but phosphorus intakes were adequate or high. Moreover, the children in South Africa responded to calcium-rich hospital diets (Thandrayon and Pettifor, 2018), whereas the Nigerian children responded more to calcium or a combination of calcium and vitamin D, and not to vitamin D alone (Thacher et al., 1999). Children with severe acute malnutrition experience hypocalcemia and many have accompanying rickets (Smilie et al., 2020).

Osteopenia and osteoporosis are conditions characterized by a reduced bone mass. In osteopenia, the T-score is low, at –1. In osteoporosis, reduction in bone mass is greater, i.e., the T-score is –2 and there is increased bone fragility and risk of fracture. The T-score is a relative rating expressed in standard deviation units describing the bone mineral density (BMD) measured at the spine, hip, or forearm. A T-score is similar to a Z-score except it is not measured against a same age person but rather against a young adult with normal BMD. Measurement is done using a variety of methods described in Section 23a.9. Osteoporosis diagnosis, according to the International Osteoporosis Foundation, (IOF), is done most often using BMD tests. For example, there is a Fracture Risk Assessment Tool for physicians to use with patients of both sexes called FRAX®. It is a web-based calculator to assess the ten-year risk of osteoporosis fracture based on specific individual risk factors, with or without BMD values (Kanis et al., 2018).There are separate country models for more than 60 individual countries, and it is available in more than 30 languages. The IOF supports the maintenance, development, and education about the use of FRAX® worldwide. The factors used in FRAX® to predict 10y fracture risk are listed in Table 23a.4.

Table 23.4: Risk factors for osteoporosis used in the FRAX® Fracture Risk Assessment Tool for physicians to assess 10-year risk of f www.iofbonehealth.org

Risk factors	Risk factors used in FRAX® to predict 10y fracture risk
Age	Increasing risk 40–90y
Sex	Female is risk factor
Weight and height	Low BMI is risk factor
Previous fracture	Previous fracture as adult is risk factor
Parent fractured hip	Indicates genetic risk
Smoking	Current use of tobacco is risk factor
Glucocorticoids	≥ 3mos prednisolone (≥ 5mg daily or equivalent) is a risk factor
Rheumatoid arthritis	Confirmed diagnosis of rheumatoid arthritis is a risk factor
Secondary causes of osteoporosis	Type I diabetes, adult osteogenesis imperfecta, untreated hyperthyroidism, hypogonadism or pre malnutrition, malabsorption, chronic liver disease
Alcohol	≥ 3 or more units of alcohol daily is a risk factor

Risk factors	Risk factors used in FRAX® to predict 10y fracture risk
Bone mineral density (BMD)	Risk increases with decreased femoral neck BMD (in g/cm^2).

Dietary patterns and lifestyle factors such as sex, body composition, genetics, and physical activity contribute to bone health (Weaver et al., 2016). Dietary factors include high intakes of protein and phosphorus (both as positive and negative influences) and suboptimal intakes of vitamin D as well as other micronutrients including vitamin K, vitamin C, magnesium, and zinc. These micronutrients are related to calcium metabolism and/or bone or connective tissue metabolism. However, based on a recent systematic review, the Level of Evidence was designated as "Inadequate Evidence" of the benefit of micronutrients (with the exception of calcium and vitamin D) on bone during acquisition of peak bone mass (Weaver et al., 2016). In examining dietary patterns, Moderate Evidence was found to support a beneficial role for dairy products on bone, whereas for dietary fiber and fruit and vegetable intake the evidence on bone was deemed "Limited". Likewise, the evidence for a detrimental effect of cola and caffeinated beverages on bone acquisition was also designated as "Limited". For bone loss in older adults, a Canadian study concluded that the "Prudent" dietary pattern consisting of fruit, vegetables, whole grains, fish, and legumes reduced bone turnover compared to the Western diet of soft drinks, potato chips, french fries, meats, and desserts (Langsetmo et al., 2016).

23a.6.3 Non-bone effects of calcium deficiency

Calcium deficiency has been implicated in effects on health that are not bone-related. These include: reduction in hypertensive disorders of pregnancy, reduction in high blood pressure, reduction in risk of colorectal adenomas, and improvements in blood cholesterol in those at risk (Cormick and Belizan, 2019). Further, calcium has been implicated in weight management (Illich et al., 2009). The evidence for or against these effects is presented below. None of these non-bone effects were used to set Nutrient Reference Values (NRVs), yet each provides additional reasons for being concerned about calcium adequacy. Blood pressure and calcium intake are inversely related in studies in humans as well as animal models, suggesting a role for calcium in maintaining normal blood pressure (Cormick and Belizan, 2019). Intracellular calcium regulates blood pressure in vascular smooth muscle cells, directly through vasoconstriction and indirectly through vascular volume control (Villa-Etchegoyen et al., 2019). Hypertensive disorders of pregnancy, particularly eclampsia and pre-eclampsia, are a significant cause of severe morbidity, long-term disability and death among both mothers and their babies (WHO, 2018). Using moderate-certainty evidence from intervention studies, WHO has made the following recommendation: In populations with low dietary calcium intake, daily calcium supplementation (1.5–2.0g oral elemental calcium) is recommended for pregnant women to reduce the risk of pre-eclampsia. Additional considerations included in this recommendation are that all of the calcium may be from foods, if possible; that the intake should be obtained in divided doses, not all at once; and that iron supplements should be taken at a different time from calcium supplements.

Reduction in risk of colorectal adenomas with dietary calcium was first postulated in 1984, whereby in the presence of calcium ions, the toxic effects of free fatty acids on colon epithelial cells could be reduced by conversion to insoluble calcium soaps (Newmark et al., 1984). Since that time, numerous studies have refined this hypothesis, and protective effects have been noted in animal studies. A systematic review and meta-analysis of randomized controlled trials regarding calcium supplementation for the prevention of colorectal adenomas found a modest protective effect of calcium (1200–2000mg doses of supplemental calcium) in prevention of adenomas (Relative Risk (RR) = 0.89, 95% CI: 0.82–0.96) (Bonovas et al., 2016).

Blood cholesterol levels may be influenced by calcium intake. A systematic review reported that calcium supplementation reduced LDL cholesterol and increased HDL cholesterol through mechanisms involving suppression of calcium-regulating hormones (e.g., PTH) that subsequently reduced intracellular calcium in adipocytes, thus stimulating lipogenesis and lipid storage (Cormick and Belizan, 2019). However, as described below in Section 23a.7, excess calcium intake may be implicated in promoting cardiovascular disease. Hence, to date there is no conclusion regarding the role of calcium intake in reducing heart disease risk through cholesterol metabolism. Body weight control by calcium has been debated for many decades. The initial suggestion arose from the finding that dairy intake facilitated weight loss better than low calorie diets without dairy, an effect that was attributed to calcium (Zemel et al., 2005). A recent trial compared the effect of supplements and dairy-containing foods and concluded that consuming 4 to 5 servings of low-fat dairy foods per day or taking calcium and vitamin D supplements in persons with borderline dietary intakes of these two nutrients is beneficial for weight loss and/or maintenance of weight loss in postmenopausal women (Ilich et al., 2009).

This page titled 23.6: Calcium deficiency (23a.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.7: Effects of high intakes of calcium (23a.7)

There are several health concerns regarding excess calcium intake over long periods of time. The risk of renal stones, risk of cardiovascular disease, and inhibition of iron absorption are current concerns (Cormick and Belizan, 2019). Previously, recognition that high intakes of calcium caused calcification of soft tissue leading to “milk-alkali syndrome” (Whiting and Wood, 1995), was the rationale for setting the first Tolerable Upper Intake Level (UL) for calcium at 2500mg/d for all persons older than 1y, including pregnant and lactating women (IOM, 1997). More recently, the U.S. Food and Nutrition Board has set the Tolerable Upper Intake Level (UL) at 2000mg/d for adults over 50y due to reports of the increased incidence of kidney stones occurring in calcium supplementation trials (IOM, 2011). A higher UL is set for other age groups in the United States and Canada as the risk of stone occurrence in younger ages is much lower. The European Food Safety Authority (EFSA) however, as noted earlier, ruled that there was not sufficient data to change from the previous UL of 2500mg derived as the amount showing no adverse effects with chronic intake (EFSA, 2012).

One concern that remains controversial is the effect of excess calcium on risk of cardiovascular disease (CVD). Many calcium trials were re-analyzed for heart effects even though none were specifically designed to assess this effect. The most recent systematic review and meta-analysis of cohort studies and randomized controlled trials (Yang et al., 2020) found that calcium intake from dietary sources did not increase the risk of CVD including coronary heart disease (CHD) and stroke. However, there was evidence that calcium supplements might raise CHD risk, especially myocardial infarction. It should be noted that many of these trials provided subjects (most often older women) with daily calcium supplements of 1000–1500mg without measuring dietary calcium intakes or accounting for the habitual calcium supplement use by subjects. Hence, calcium intakes may have exceeded 2000mg/d, the UL level set for adults over 50y by the Institute of Medicine (IOM, 2011).

This page titled 23.7: Effects of high intakes of calcium (23a.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.8: Biochemical indices of calcium status (23a.8)

As calcium homeostasis is tightly regulated, there is no satisfactory single test to assess calcium status directly on a routine basis. Calcium is an important extracellular electrolyte (Ca^{++}) and can be measured in serum or plasma. Normally values are reported as total calcium concentration, but ionized calcium can be an important measurement. Hypocalcemia and hypercalcemia (reduced and elevated serum calcium concentrations, respectively) are not likely to be caused by dietary intakes alone except on a transient basis. Urinary calcium excretion (calciuria) may give some clues as to calcium metabolism but as levels are equally likely to be from bone turnover or from dietary intakes, calciuria is difficult to interpret on its own. Instead, biochemical markers of bone formation and bone resorption are used as measures of calcium status as they reflect need for calcium (shown as excess bone resorption) and calcium adequacy (shown as appropriate bone formation) (Bonjour et al., 2014).

23a.8.1 Serum calcium and its regulation

Serum calcium is not a test of nutritional status but rather reflects hormonal regulation of calcium, not calcium balance. Calcium is present in the blood in three different forms: as free Ca^{2+} ions, bound to protein (about 45%), and complexed to citrate, phosphate, sulfate and carbonate (about 10%). Calcium in the blood (and in extracellular fluid) is kept constant at 2.5mmol/L (range 2.25–2.6mmol/L), but it is ionized calcium (between1.1–1.4mmol/L) that is the form that is actively transported across membranes (EFSA, 2012). Values decrease with age in men. Females usually have slightly lower concentrations (2.20–2.54mmol/L), associated with small differences in serum albumin content and the calcium-reducing effect of estrogens. Serum calcium decreases by 5–10% up to the end of the third trimester of pregnancy, after which concentrations rise. This decrease in total serum calcium may not be “true” hypocalcemia, as it is not due to a decrease in ionized calcium. However, changes in the serum chemistries and calcium-regulating hormones can be mistaken as a disorder of calcium (Almaghamsi et al., 2018).

Hypocalcemia (low serum calcium) and hypercalcemia (elevated serum calcium) indicate serious disruption of calcium homeostasis but generally do not reflect calcium balance. In severe hypocalcemia, blood calcium is < 2.12mmol/L, and in severe hypercalcemia blood calcium is > 2.72mmol/L. Serum calcium levels must be corrected for the serum albumin concentration before confirming the diagnosis of hypercalcemia or hypocalcemia (Fong and Khan, 2012).

Hypocalcemia is most commonly a consequence of vitamin D inadequacy or hypoparathyroidism, or a resistance to either of these hormones. Hypocalcemia has also been associated with the use of certain drugs including bisphosphonates, cisplatin, anti-epileptics, diuretics, and proton pump inhibitors. Symptoms of hypocalcemia may include muscle spasms, cramps, tetany, numbness, and seizures, but also can be asymptomatic (Fong and Khan, 2012). Hypocalcemia is common in the critically ill patient but the interpretation of the low serum calcium level is complicated by the co-existence of hypoalbuminemia and disorders of acid-base balance, thus in these patients hypocalcemia is defined in terms of ionized calcium concentration rather than total calcium concentration. Mechanisms causing hypocalcemia in critically ill patients include: hypoparathyroidism, severe vitamin D deficiency, hyperphosphatemia, infusion of chelating agents such as citrate, alkalosis caused by hyperventilation, and pancreatitis that causes a rise of free fatty acids (FFAs) that result in increased binding of calcium to albumin (Kelly and Levine, 2013).

Hypercalcemia occurs in association with hyperparathyroidism, hyperthyroidism, and sarcoidosis and when large parts of the body are immobilized (e.g., when a patient is hospitalized after spinal cord injury or major limb-bone fractures). In the latter cases, calcium from the rapidly atrophying bone is released into the circulating body fluids. In addition, hypercalcemia is one of the characteristic features of vitamin D intoxication. It arises from hyperabsorption of intestinal calcium and to a lesser degree from the release of calcium from bone (Peacock, 2010). Severe chronic hypercalcemia can result in nephrolithiasis (renal stones) and impairment of kidney function, as well as in calcification of soft tissues (e.g., nephrocalcinosis and vascular calcification), the latter when phosphorus concentrations in the blood are also high, as in renal insufficiency (EFSA, 2012).

Hypercalcemia may occur in advanced stages of the following cancers: multiple myeloma, breast cancer, parathyroid cancer, lung cancer, kidney cancer, lymphoma, leukemia, and bone metastases (Canadian Cancer Society, 2020). This hypercalcemia causes many unpleasant clinical symptoms such as vomiting, anorexia, constipation, frequent and increased urination, a patient may seek medical treatment at which time blood levels may be measured along with urinalysis and kidney function tests. Treatment would involve initially reducing hypercalcemia through fluid replacement followed by use of calcium-losing diuretics such as furosemide. As well, anti-resorptive medications such as bisphosphonates and denosumab, that reduce bone turnover, may be used to prevent calcium release from bone resorption.

23a.8.2 Measurement of serum calcium

Serum calcium levels have traditionally been determined by flame atomic absorption spectrophotometry (AAS) (Zettner and Seligson, 1964) or by automated procedures that use the o-cresolphthalein complexone method (Bourke and Delaney, 1993).

Earlier methods used flame photometry. For surveying the American population, the National Health and Nutrition Examination Survey (NHANES) uses a DxC800 system with ion-selective electrode methodology to measure calcium concentrations in serum, plasma, or urine. Only NHANES III provides ionized calcium values for the American population. Considerable differences in the quality of serum calcium measurements among laboratories may occur, emphasizing the necessity for rigorous quality-control practices. In general, the coefficient of variation for the AAS method is less than that for flame photometry (Linko et al., 1998). Analytical variation for serum calcium via AAS ranges from 0.9%–3.04% (Gallagher et al., 1989).

As such a large amount of calcium in serum is bound to albumin, it follows that concentrations can be influenced by changes in serum albumin levels. As a result, when serum albumin levels are outside the normal limits (Iqbal et al., 1988), as may occur in liver disease or malnutrition, serum calcium values should be corrected, as noted earlier. The correction is approximate, and if there is uncertainty in the interpretation of the results, serum ionized calcium should be directly determined.

In many methods. serum rather than plasma should be used for calcium analysis as the former has no anticoagulant added. Most anticoagulants, with the notable exception of lithium heparin, interfere with the determination of calcium: they complex with or precipitate the calcium in the sample.

23a.8.3 Measurement of serum ionized calcium

Fasting blood samples are recommended for the measurement of ionized calcium in serum, to eliminate the effect of a recent meal. The samples should be stored anerobically at 0°C prior to the assay. Serum ionized calcium can be measured using commercially available ion-specific electrodes (Wandrup and Kvetny, 1985). Care must be taken to ensure that the calcium standards used for the assay contain sodium and chloride at the same levels as in the test samples. Further details on preparing blood for ionized calcium are published (Hamroun et al., 2020). Factors affecting serum ionized calcium levels include physical activity and circadian rhythm.

Many clinicians estimate ionized calcium using formulas that “adjust” total calcium with albumin and/or protein levels, and sometimes including other blood constituents such as phosphate (Hamroun et al., 2020). There are over 30 published formulas, with the most widely used one being the Payne Formula:

$$\begin{gathered}\text{Adjusted calcium (mmol/L)} = \\ \text{Totalcalcium (mmol/L)} + 0.02(40 - \text{ serum albumin (g/L)})\end{gathered}$$

However, caution is advised before applying this formula as it is not applicable in many situations.

The clinical utility of ionized calcium ranges from conditions with moderate evidence (drug-induced calcium disorders, cardiovascular outcomes), good evidence (post-thyroidectomy hypocalcemia), and very good to strong evidence (calcium disorders in critically ill patients, primary hyperparathyroidism, malignancy-related hypercalcemia). There is poor to fair evidence for use of ionized calcium in pre-eclampsia (Hamroun et al., 2020).

Interpretive criteria

The concentration of serum ionized calcium is tightly maintained within a physiologic range of 4.4–5.4mg/dl (1.10–1.35mmol/L) (Peacock, 2010).Reported levels of serum ionized calcium in adults vary. In NHANES III, mean age-adjusted serum ionized calcium values were presented for male and female non-Hispanic Caucasians, African Americans, and Mexican Americans for two age groups: 25–59y and 60–89y; the reported mean age-adjusted value for serum ionized calcium was 1.237mmol/L for males and 1.232mmol/L for females (Vargas et al., 1998). Mean age-adjusted values for African Americans were slightly higher.

23a.8.4 Calcium balance

Calcium is lost daily through urinary and fecal excretion. In the gastrointestinal tract, calcium is lost through sloughing of cells as well as unabsorbed calcium in the feces. A small amount of calcium is lost in sweat. Replacement of these losses must occur through the diet (Weaver and Peacock, 2019). A zero calcium balance means all losses have been replaced, and there is no net

bone loss to provide serum calcium during periods of fasting. A positive calcium balance indicates that losses are less than intake, and implies calcium accretion into bones, an expected finding during growth and during pregnancy. A negative calcium balance means losses exceed intake, and the additional calcium lost is from bone. Over time a negative calcium balance indicates bone formation is less than bone resorption. Table 23a.5 provides a summary of these assumptions.

Table 23a.5: Calcium balance, bone, and the life stages.

Ca Balance	Bone	Life Stage
Positive	Accretion Formation > Resorption	Childhood Adolescence Pregnancy
Neutral (or zero)	Maintenance Formation = Resorption	Adults
Negative	Loss Formation < Resorption	Post-menopausal Women Older Adults

Measurement of calcium balance was a mainstay of nutrition research prior to the introduction of bone imaging equipment. These early experiments required facilities (metabolic units) where participants could receive all meals, sometimes for several months. A typical experimental protocol would adhere to the following steps:

1. participants ingest the experimental diet for at least one week adaptation;
2. in each 15d cycle participants would consume only those foods and beverages provided, and every day collect 24h urine samples;
3. additionally participants would collect all feces along with consumption of a fecal marker during these 15d;
4. study research personal prepare composites of food, urine, and feces representing each of three 5d cycles; and
5. calcium was measured in the food and fecal composites after dry ashing to remove the organic material, followed by quantitative dilution, while urine was acidified to prevent precipitation of calcium (Anand and Linkswiler, 1974).

Balance is calculated as Ca-intake minus Ca-losses. Unless participants resided in a metabolic unit, calcium losses through sweat were not taken into consideration. Overall, this method is expensive and requires exceptional participant cooperation.

23a.8.5 Urinary calcium as a biomarker

The amount of urinary calcium excreted is highly variable as it is dependent on the diet and the efficiency of intestinal absorption. Under conditions of bone accretion, one expects less calcium to be excreted as bones are forming, while during bone loss, more excretion might be observed. Measure of 24h urine calcium may provide some indication of calcium intake as extremely low excretion rates may indicate low calcium intake. However, extremely high excretion rates, (i.e., hypercalciuria), may be indicative of bone loss or clinical conditions unrelated to dietary intake.

For calcium balance, a 24h collection of urine is necessary to obtain the estimate of calcium excretion, yet this procedure requires enormous participant compliance. To gauge an indication of urinary calcium rate, some studies measure calcium in spot urine samples, usually after an overnight fast, when urinary calcium is expressed per mg or mmol of urinary creatinine. Using urinary creatinine in the denominator corrects for variability of urinary excretion as the amount of creatinine excreted in 24 hours is constant for individuals. Further, creatinine excretion can be used as an measure of muscle mass as it is made from creatine. Hence, when expressing urinary calcium excretion as a ratio of urinary creatinine excretion, one can equalize excretion between participants of dissimilar body mass. There is good correlation between excretion ratios of calcium expressed per mg or mmol of urinary creatinine in 24h urine collections and those based on spot urine collections (Ilich et al., 2009).

23a.8.6 Biomarkers of bone formation and resorption

Bone remodeling is an important process that continuously renews bone throughout life. It repairs micro-damage, maintains mineral homeostasis, and ensures mechanical proficiency by modifying the micro-architecture. Bone remodeling is regulated by a variety of systemic and local factors. Bone remodeling involves the following cycle of events: osteocyte signaling, recruitment of osteoclasts to begin resorption, degradation and removal of bone, reversal, formation of new bone by osteoblasts, and then a period of resting (Bonjour et al., 2014). Each of the stages in bone remodeling can produce molecules, most of which are pro-

tein-derived, that enter the plasma. These biochemical markers of bone remodeling (see Table 23a.6) can be used to measure subtle changes in the rate of formation or degradation of the bone matrix brought about by dietary influences, such as availability of calcium, phosphate, and protein. When remodeling rates are changing, a combination of biomarkers, such as one formation marker and one resorption marker can provide more information than a single marker (Watts, 1999). In addition, they can provide an earlier estimate on the effect of dietary interventions or drug treatments compared to measures of BMD. However, because of increases in bone turnover during the night, bone biomarker concentrations are affected by circadian rhythm. For all the bone biomarkers shown in Table 23a.6, their measurement is generally recommended using automated platforms which have improved technical performance (Bonjour et al., 2014).

Table 23.6 Proteins or protein fragments released from osteoclasts measure bone resorption while those from osteoblasts measure Bonjour et al., Nutrition Research Reviews, 2014, 27:252–267.

Bone Resorption (Osteoclast activity)
Osteoclast number • Tartrate-resistant acid phosphatase (TRAP)
Bone matrix absorption • Pyridium cross-links (PYR) • Carboxy terminal telopeptide (CTX) • terminal propeptide (PINP)
Bone Formation (Osteoblast activity)
Osteoblast protein synthesis • Procollagen type
Osteoblast activity • Bone alkaline phosphatase (BALP) • Osteocalcin (OC)

A description of the commonly used bone biomarkers shown in Table 23a.6 is outlined below. Of these, use of only two, namely P1NP (for formation) and CTX-1 (for resorption) are recommended by the International Osteoporosis Foundation (IOF), and the International Federation of Clinical Chemistry (IFCC) because they are well characterized for use in fracture risk prediction and for monitoring osteoporosis treatment (Kim et al., 2020).

23a.8.6.1 Biomarkers of bone formation

Bone formation is determined by measuring proteins that increase in serum as a result of osteoblast activity during bone formation. These include bone alkaline phosphatase (BALP), which is measured as enzymatic activity, and the vitamin K-dependent protein osteocalcin (OC). A third marker, procollagen type I N-terminal propeptide (PINP) that reflects the protein synthesis capacity of the osteoblast can also be used. Early measurement of these three bone formation markers (OC, BALP and particularly PINP) has shown positive correlations with subsequent bone mineral density improvements in trials of teriparatide, a drug promoting bone formation (Bonjour et al., 2014). Bone alkaline phosphatase (BALP), an isoenzyme also known as bone-specific alkaline phosphatase or skeletal alkaline phosphatase, is measured in serum, and is essential for normal bone mineralization. Several techniques have been developed to measure the activity of this enzyme. In adults approximately 50% of serum total alkaline phosphatase is present as bone alkaline phosphatase, whereas during childhood, the bone isoenzyme predominates. BALP may also be measured when assessing vitamin D deficiency (see chapter 18b.12.2) during which osteoblast activity is suppressed due to low calcium absorption. Thus, BALP is not specific to calcium deficiency and interpretation of its activity must be accompanied by other measurements.

Osteocalcin (OC), measured in serum, is also called a bone γ-carboxyglutamic acid–containing protein because it contains up to three γ-carboxyglutamic acid residues. This small non-collagenous protein is specific for bone tissue and dentin and comprises about 1–2% of total bone protein. During synthesis of osteocalcin, a small fraction is released directly into the blood. Osteocalcin is cleared by the kidneys so that serum levels depend on renal function. The half-life of serum osteocalcin is short (15–70min). Uncertainties persist as to the exact role of osteocalcin. Indeed, osteocalcin is secreted solely by osteoblasts, yet has only minor effects on bone mineralization. Instead, osteocalcin acts as a hormone-like factor in glucose homeostasis, brain development, cognition, and male fertility (Moser and van der Eerden, 2019).

Procollagen type I N-terminal propeptide (PINP) measured in serum, is the preferred marker to measure bone formation (Bonjour et al., 2014). Most of the organic matrix of bone is type I collagen associated with non-collagenous proteins, and Type I collagen is rich in the amino acid hydroxy-proline. During the conversion of procollagen to collagen, which occurs extracellularly, the pro-collagen propeptides, procollagen type I carboxy- terminal propeptide (PICP) and procollagen type I aminoterminal propeptide (PINP) are released into the circulation. Although both bone alkaline phosphatase and PINP are markers of bone formation, PINP is the reference biomarker for bone formation (Kim et al., 2020).

23a.8.4.2 Biomarkers of bone resorption

Bone resorption markers include the pyridinium cross-links (PYR) subdivided into pyridinoline and deoxypyridinoline together with the associated telopeptides, i.e., carboxy terminal telopeptide (CTX) and N-terminal telopeptide (NTX), all of which are released during collagen breakdown (Table 23a.6). A second marker of bone resorption is tartrate-resistant acid phosphatase (TRAP), a measure of osteoclast number (Bonjour et al., 2014).

Pyridinium cross-links (PYD) measured in urine were among the first bone biomarkers of resorption, but are rarely measured today. The crosslinks are hydroxylysylpyridinoline (known as pyridinoline, PYD) and lysylpyridinoline (known as deoxypyridinoline, DPD) and are important for the structural integrity of the collagen. Pyridinoline is widely distributed in both type I collagen of bone and type II collagen of cartilage, as well as in smaller amounts in the other connective tissues, and is thus considered non-specific. In contrast, deoxypyridinoline is found almost exclusively in type I collagen of bone(Kuo and Chen, 2017). Both the cross-links are released into the circulation after mature tissue collagen is degraded. They are not re-utilized or metabolized in the liver but, instead, are excreted in the urine unchanged, in peptide-bound and free forms. Neither PYD or DPD are commonly measured (Bonjour et al., 2014).

Cross-linked N-telopeptides (NTX) and C-telopeptides (CTX) measured in serum or urine are now the most commonly measured bone biomarkers of resorption. C-telopeptides (CTX) are better characterized than NTX and are the preferred bone resorption marker, although both have similar properties and origins. CTX is the reference marker for bone resorption (Kim et al., 2020). CTX and NTX are carboxy-terminal and amino-terminal fragments of collagen, respectively which are released into blood and subsequently excreted in the urine with cross-links attached (Bonjour et al., 2014). However, serum CTX levels are decreased by food intake so blood withdrawal must take place in the fasting state (Kuo and Chen, 2017). Measurement of biomarkers can be conducted using more than one detection method for each biomarker. A complete list of all the biomarkers discussed above as well as others still under consideration is available (Kuo and Chen, 2017), combining the biomarkers to generate a biomarker risk score is under development to enhance the accuracy of predicting fractures (Kim et al., 2020).

This page titled 23.8: Biochemical indices of calcium status (23a.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

23.9: Measurement of bone mineral content and bone mineral density (23a.9)

Noninvasive techniques can be used to provide an indirect assessment of calcium balance as well as calcium “stores” in bone by measuring the bone mineral content (BMC) or bone mineral density (BMD). These methods include dual X-ray absorptiometry (DXA, also called DEXA) and quantitative computed tomography (QCT), both major diagnostic tools for osteoporosis. Ultrasound (US), although less accurate for medical diagnosis, can serve as a tool for screening and for field work. These technologies measure the mineral content of bone in the appendicular skeleton, the axial skeleton, or the total skeleton, measurements that relate directly to the calcium content of the skeleton. As almost all of the calcium in the body is stored in the skeleton, it follows that skeletal mineral content can be used as an indirect measure of body calcium stores. These same methods can also be used to monitor the response to changes in calcium intakes over relatively long time periods, i.e., months or even years.

Over the past 40 years technologies such as DXA and QCT have been developed and refined to measure calcium in hard tissues, and thus estimate of total body mineral content, which in turn can be used to calculate total body calcium. These technologies are also used to determine bone integrity for the diagnosis of osteoporosis and other bone diseases. They can also be used during treatment to evaluate efficacy as well as in research studies for the development of new drugs, exercise regimens, or nutrition interventions (IOF, 2020).

The measurement of bone mineral using DXA is usually expressed as the bone mineral density, which corrects for projected bone area. However, the measurement is not a true density measurement, but instead, a measure of areal density (g/cm^2) derived by dividing bone mineral content (g) by the scanned bone area. Bone mineral density is calculated in this way to reduce the biological variation observed in values for bone mineral content at all ages and thus to increase the statistical power of detecting abnormal values. Nevertheless, bone mineral density does not adequately correct for bone area and body size and thus must be interpreted carefully (Prentice et al., 1994). In contrast, QCT uses multiple slices to construct three-dimensional images, thus providing a true volumetric measure for BMD (Adams, 2009).

Diagnosis for osteoporosis is based on the difference between the measured bone mineral density (BMD) of the participant and the mean BMD of healthy young adults, matched for sex and ethnicity. The difference is expressed relative to the young adult population standard deviation, and reported as a T-score. WHO, 1994 has defined both osteoporosis and osteopenia based on T-score values. An individual with a T-score less than –2.5 at the spine, hip, or forearm is classified as having osteoporosis; a T-score between –2.5 and –1 is classified as osteopenia; and a T-score greater than –1 is healthy. These WHO definitions should not be applied to other bone mineral density measurement sites.

The International Society for Clinical Densitometry (ISCD) has established guidelines for bone density testing using DXA. In addition, they provide specific guidelines for diagnosing osteoporosis based on T-score values for postmenopausal women, men (20y and older), pre-menopausal women (20y to menopause), and children (males or females < 20y). Persons using DXA technology should be certified, and facilities should be accredited; consult the ISCD for advice on certification and accreditation.

23a.9.1 Single-photon absorptiometry

Single-photon absorptiometry was the first practical noninvasive method that was developed to examine the peripheral skeleton (Wahner et al., 1983).although now it has mostly been replaced by DXA (IOF, 2020). The original method was based on the assumption that the bone mineral content is inversely proportional to the amount of photon energy transmitted through the bone under study. A mono-energetic photon source is used, usually ^{125}I or ^{241}Am. The technique is fast, taking approximately 5min. With this method, bone mineral density is measured with a precision (%CV) of less than 2% and with an effective dose to the participant of only 0.1μSv. The site most frequently selected for measurements by single-photon absorptiometry is the lower radius, at approximately one-third of the distance from the styloid process to the olecranon. Single-photon absorptiometry is not suitable for measurements on the axial skeleton because the technique requires that a uniform thickness of soft tissue surrounds the bone (Neer, 1992). Newer technologies have replaced SPA for human studies as SPA cannot measure those sites (hip or spine) needed for clinical diagnoses. In the past four decades SPA machines have been adapted and used for animal studies of bone

health (Sequeira et al, 2020).

23a.9.2 Dual X-ray absorptiometry (DXA)

The first commercial DXA scanner was marketed in 1987. There are now several commercial versions, each using similar measurement procedures The original DXA method used an isotopic source, typically ^{153}Gd , that emitted two low-energy gamma rays at 44Kev and 100KeV. The short half-life of the isotopic source ^{153}Gd, however, limited the precision with which long-term changes in bone mineral content in the same participant could be made using this technique. This limitation led to the radioactive ^{153}Gd source being replaced with an X-ray tube behind a K-edge filter. The filter converts the polychromatic X-ray beam into one with two main energy peaks at 40Kev and 70KeV. These two congruent X-ray beams are passed through the body, and the ratio of beam attenuation at the lower energy relative to that at the higher energy differentiates between bone mineral, bone-free mass, and the fat mass (Kim et al., 2018). As a result, this modified instrument is termed a dual energy X-ray absorptiometer, abbreviated as DXA (or DEXA).

The DXA instrument is used not only for bone mineral measures but also to assess body composition, as described in Chapter 14. Only the measurements of bone mineral mass (g), bone mineral content (g/cm), and "areal" bone mineral density (g/cm^2) are discussed here. Using DXA, measurements of areal bone mass density (aBMD) can be made at the lumbar spine, femur, and forearm, as well as total body. Measurements of body mineral density at the lumbar vertebrae (L2 to L4) are often favored because they are sensitive to the changes associated with aging, disease, and therapy. Care must be taken when positioning the participant for the scan to ensure that precise and accurate data are obtained. Under some circumstances, it may be appropriate to make a whole-body scan. This is quick, taking from 3min to 35min, depending on the instrument used and the age of the participant, and requires very little cooperation from the participant. As a result, scans can be readily performed on young children, the elderly, and even persons who are sick, although pregnant women should be excluded (Kim et al., 2018).

Dual-energy X-ray absorptiometry is now the dominant method for measuring bone loss for clinical diagnosis of osteopenia and osteoporosis. In intervention studies care must be taken to ensure that both the length of the intervention and the size of the study population are sufficient to allow the detection of small changes in bone mineral density. A meta-analysis of trials reported that increasing calcium intake from dietary sources increased BMD by 0.6–1.0% at the total hip and total body at one year, and by 0.7–1.8% at these sites and the lumbar spine and femoral neck at two years (Tai et al., 2015). Larger decreases in bone loss have been observed after treatment with estrogen, calcitonin, or biphosphates.

When comparing measurements using DXA based on different instruments and technologies, it is important to acknowledge the potential for systematic differences arising from methodological factors which could mask those associated with biological variation. Methodological factors may include differences in instrument calibration methods, whereas variations in body thickness, distribution of body fat, and the fat content of the bone marrow are major sources of biological variation. Of the latter, DXA is especially sensitive to discrepancies in body thickness, which can lead to systematic differences in the bone mineral estimates on thin and obese subjects. In a global survey of fracture liaison services, although access to DXA was reported to meet the needs (Clynes et al., 2020), only around 50% of institutions confirmed adherence to basic DXA quality and reporting procedures, and many required educating the operators/interpreters. Indeed, significant variability worldwide in both the quality and access to DXA services was reported in this survey.

Calcium balance using DXA may be determined by obtaining a total body (TB) DXA scan that provides bone mineral content (BMC). Scans of total body bone mineral content (TBBMC) can be used to calculate calcium requirements of specific life cycle groups, on the assumption that bone mineral is 32·2% calcium (Vatanparast et al., 2010). Use of this value is a better approximation of calcium in bone than simply dividing the atomic weight of calcium (40g/mol) by the molecular weight (i.e., 1005g/mol) of hydroxyapatite $Ca_{10}(PO_4)_6(OH)_2$ because the mineral in bone is hydrated so water molecules contribute to the weight of bone mineral content.

To derive the Estimated Average Requirement (EAR) for calcium for adolescents 9–18y, the U.S Food and Nutrition Board applied the factorial approach to estimate the physiological calcium requirements (IOM, 2011). Estimates of calcium losses via urine, feces, and sweat were derived from metabolic balance studies to which was added the calcium accrued as bone mineral for growth (derived from TBBMC). The latter was measured in a cohort of Canadian males and females in two age groups, 9–13y and 14–19y as shown in Table 23a.7 (Vatanparast et al., 2010). To translate the physiological requirements for absorbed calcium

(from the factorial approach) into the EAR, it is necessary to take into account the proportion of calcium in the habitual diet that is absorbed by the intestine. Assuming an efficiency of absorption of 38% from an omnivorous diet, the equation used for the estimated Ca dietary requirement (i.e., the EAR) using the formula for adolescence which is a period of rapid growth:
EAR = Ca losses + (Ca needs for growth / 38%)
(IOM, 2011).

Table 23.7. Determining calcium requirements during adolescence (age 9–18y) using total body bone mineral content (TBBMC). Source: Vatanparast et al., (2009). British Journal of Nutrition 103: 575–580.

Factorial Components	Girls 9–18y (mg/d)	Boys 9–18y (mg/d)
Ca accretion from TBBMC	121	175
Urinary Ca Losses	106	127
Endogenous Fecal Ca	112	108
Sweat Ca Losses	55	55
Total Ca physiological need	394	465
Absorption (%)	38	38
Dietary Ca requirement (EAR)	1037	1224

23a.9.3 Computerized tomography (CT) and peripheral quantitative CT (pQCT)

Computerized tomography (CT), often called quantitative CT (QCT) when referring to bone measurements, can be used to measure bone mass of both the appendicular and axial skeleton, although the method is not feasible for population studies. The advantage of QCT is that a true three-dimensional (i.e., volumetric) bone density can be obtained, so results can be expressed in g/cm^3 rather than the two-dimensional areal density measured by DXA. Bone can also be identified as cortical or trabecular in nature on the basis of bone density measurements. This is an advantage because some forms of osteoporosis are predominantly trabecular in character. The equipment is not portable, and the radiation dose required per slice for imaging is relatively high (50–500μSv), although it can be reduced if scans are limited to regions of specific interest. The cost of the equipment and its maintenance is high.

The use of QCT technology applied to specific bone sites yields information about microarchitecture. Methods that can be used include micro-CT and peripheral Quantitative CT (pQCT), with the latter having widespread use. The equipment for pQCT is small relative to DXA and CT, and provides imaging of both the arm at the distal radius and the lower leg (tibia) sites to assess bone microarchitecture. High-Resolution pQCT (HR-pQCT) has been developed to improve the resolution problems of pQCT. Recently, second-generation HrpQCT has been introduced which differs from the first-generation HR-pQCT in scan region, resolution, and morphological measurement techniques (Whittier et al., 2020).

23a.9.4 Quantitative ultrasound

Quantitative ultrasound (QUS) is used for measurements of the bone density of the peripheral skeleton such as the proximal arm (radius), lower leg (tibia) or heel. A wide variety of equipment is now available to measure the attenuation of a sonographic pulse as it passes through bone and is scattered and absorbed by the trabeculae. The heel is the most frequently used measurement site because it has a large volume of trabecular bone that is accessible for transmission measurements. However, there are ultrasound machines on the market that measure multiple skeletal sites. Quantitative Ultrasound has the advantage of being radiation-free and a portable system.

Quantitative ultrasound measures the speed of sound (SoS) in meters per second (m/s). In a study of different age groups using QUS, measures of SoS (m/s) varied according to life stages that reflected bone development (Rivas-Ruiz et al., 2015). For example,

bone SoS accretion was found to begin 5y earlier in girls than boys ($p < 0.05$), a sex-related difference that mirrors results seen in DXA scans (Vatanparast et al., 2010). The maximal (peak bone) SoS was noted at 28y at the radius site and at 22y at the tibia site. Women who were postmenopausal (age 45–50y) showed a significant decrease in SoS measures at both these sites ($p < 0.05$) compared with men, as would be expected. Thus, while the precision of QUS is poor relative to DXA, the equipment may be suitable for characterizing bone in populations where DXA is not a feasible option. However, QUS has no clinical applications and is not used to define T-scores for diagnosing osteoporosis or monitoring treatment. However, a systematic review has shown that quantitative ultrasound is an independent predictor of fracture for men and women when QUS values are low (McCloskey et al., 2015).

This page titled 23.9: Measurement of bone mineral content and bone mineral density (23a.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24: PHOSPHORUS (CHAPTER 23B)

Abstract

The essential nutrient phosphorus (in nature occurring as phosphate) is ubiquitous in all the foods we eat, the human body and, in effect, all living organisms. phosphate is critical to structural and biochemical functions needed to secure energy, reproduce and grow. Most of the body's phosphate is contained within bones, teeth, membranes and intracellular spaces; however, it is the 1% present in the extracellular space, serum that is clinically measured to inform about physiologic and nutritional phosphate status.

Serum phosphate in healthy individuals usually reflects phosphate balance that is maintained within a narrow range by hormonal control of renal reabsorption and excretion, and intestinal absorption when dietary phosphate intakes are low or excessive. Regulation of serum phosphate involves the interplay of four organs (kidneys, intestine, bone and parathyroid glands), phosphate in these organs, and the actions of three endocrine hormones (parathyroid hormone, calcitriol (the active form of vitamin D), and bone-secreted fibroblast growth factor-23 (FGF-23), all of which influence the activity of the phosphate transporters to increase or decrease absorption, reabsorption or excretion of phosphate.

Hyperphosphatemia (i.e., serum phosphate > 1.45mmol/L) is often related to excess dietary phosphate intake by the consumption of phosphate additive-rich processed foods, or the typical Western diets when kidney function is impaired. Higher serum phosphate has been associated with disruption of endocrine pathways that may link high phosphate intake with pathology associated with chronic disease risk, including cardiovascular disease. In contrast, hypophosphatemia (i.e., serum phosphate < 0.87mmol/L) is rarely related to dietary deficiency of phosphate except in cases of severe malnutrition, and more likely due to inborn errors of metabolism or tumor production of excess FGF-23 that causes renal phosphate wasting and bone disease (rickets and osteomalacia).

24.1: Phosphorus (23b.1)

24.2: Biological forms of phosphate and their measurement (23b.2)

24.3: Interpretive criterion- serum phosphate (23b.3)

24.4: Phosphate balance (23b.4)

24.5: Hormonal regulation of serum phosphate (23b.5)

24.6: Nutrient reference values (23b.6)

24.7: Dietary sources

24.8: Abnormalities in serum phosphate (23b.8)

24.9: Inherited and tumor-induced disorders of phosphate metabolism

This page titled 24: Phosphorus (Chapter 23b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.1: Phosphorus (23b.1)

Phosphorus is the 11th most abundant element yet phosphorus is not present in nature as elemental phosphorus, instead occurring mainly bound to oxygen as phosphate (Ferro, 2018). Deposits of phosphate-rich rock in the earth's crust slowly release phosphate which builds up in soils over time, entering the food chain via soil microbes, and then crops, livestock, and other components of the human food supply. However, deposits of rocks rich in phosphate are limited across the globe. When mined, these phosphate deposits are largely used as fertilizers for crop growth (Ferro, 2018).

Phosphorus is often the key growth-limiting factor for all living things. As an essential nutrient, phosphorus functions in critical pathways and cellular components in all life forms on earth, ranging from subcellular viruses to complex plants and animals, all dependent on phosphate for energy, growth, reproduction, structure, and homeostasis facilitated through signal transduction. Phosphates participate in all biological processes providing energy stored in phosphodiester bonds of ATP (the phosphodiester backbone of RNA and DNA). Other functional roles include the structural integrity of cell membranes as phospholipids, regulation of acid base balance, mineralization of teeth and bones, lipid transport in blood, and signalling pathways essential to maintaining phosphate homeostasis. As phosphate has such a critical role in so many biological processes, phosphate homeostasis must be tightly regulated. In conditions when sources of phosphate are deficient, growth or reproduction is limited, whereas with excess, toxicity may occur, which in humans may manifest as disease (Hernando et al., 2021).

This page titled 24.1: Phosphorus (23b.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.2: Biological forms of phosphate and their measurement (23b.2)

The three forms of phosphorus bound to oxygen that occur in nature are shown in Figure23b.1: an ionic anion, an inorganic salt, and as an organic compound, using phytate as the example. Inorganic or mineral phosphate largely comprises the different salts of orthophosphate that occur in greater abundance than pyrophosphate. Within the pH range of the human body, the two main forms of orthophosphates are $H_2PO_4^{-1}$ and HPO_4^{2}. The organic form of phosphate occur when bound to a carbon atom of protein, lipids, nucleic acids, and other organic compounds, usually through phosphate ester linkages. The total body phosphate content in an adult human is about 900g of elemental phosphorus, existing mainly in the skeleton and teeth, with less amounts in soft tissues.

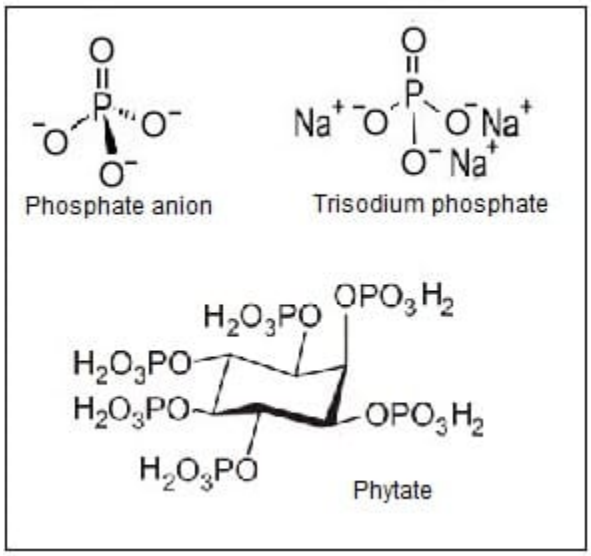


Figure23b.1 Phosphate anion, Trisodium phosphate, and Phytate — the latter redrawn from Marolt and Kolar (2020).

The phosphate in bone and teeth is present as calcium phosphate hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$). Less than 1% of phosphorus occurs in the extracellular space. Intracellular phosphate consists largely of organic molecules such as creatine phosphate, ATP, nucleic acids, phospholipids and phosphoproteins in concentrations of phosphate of about 1mmol/L.

When assayed in biological fluids and tissues, only inorganic phosphate is measured. Prior to analysis, the analytical samples other than serum must be ashed to remove the organic material, after which the residue is dissolved in dilute acid in preparation for analysis by AOAC, AAS or ICP-MS, which report total phosphorus as elemental P. In food composition databases and tables, the values are expressed as mg elemental P per serving or per 100g food item.

Clinically, the terms phosphate and phosphorus are used interchangeably. However, because elemental phosphorus does not occur in the human body, phosphorus is usually measured as mg phosphate and can be converted to molar P concentration by dividing the measured weight in mg by the atomic weight of P (31).

This page titled 24.2: Biological forms of phosphate and their measurement (23b.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.3: Interpretive criterion- serum phosphate (23b.3)

Serum phosphate is the most frequently used biomarker of phosphorus status in a clinical setting, and is usually measured in the fasting state. However, the measurement of a single fasting serum phosphate concentration represents only a small portion of the total body phosphate, and hence does not always reflect the body phosphate stores. Measurement of serum or plasma phosphate concentration requires the use of anticoagulants such as heparin which do not interfere with the color reaction described for the AOAC spectrophotometric method described in 23b.2. Hemolyzed samples are not suitable for phosphate measurement as erythrocyte phosphate confounds the measurement and hemoglobin contributes color interference. Serum phosphate concentrations can also be affected temporarily by acute shifts of phosphate between intracellular and extracellular compartments without affecting total body content (Uribarri & Oh, 2018).

Table 23b.1 Normal Inorganic Serum phosphate Values for Children and Adults. For more information on disorders of serum phosphate see (Koumakis et al., 2021).

Sample	Reference Ranges	
	mg/dL	mmol/L
Cord	3.7 - 8.1	1.2 - 2.8
Child	4.5 - 5.5	1.45 - 1.78
Adult	2.7 - 4.5	0.87 - 1.45
Older Adult > 60y	M: 2.3 - 3.7 F: 2.8 - 4.1	M: 0.74 - 1.2 F: 0.90 - 1.3

Serum phosphate concentrations are maintained within a narrow range (see Section 23b.5 for details of the hormonal regulation of serum phosphate). In adults, total serum inorganic phosphate ranges between 0.87-1.45mmol/L (Table 23b.1): 56% is ionized, 20% bound to protein and 24% bound to other cations. However, there is a significant amount of organic phosphate in serum (7.5-8.0mg/dL or 2.4-2.6mmol/L) which is not included in the analytical method used by clinical laboratories. When serum phosphate concentrations fall below the normal range, a condition called hypophosphatemia occurs, whereas for concentrations above the normal range, hyperphosphatemia develops; serious clinical consequences can arise from both conditions.

Several factors affect serum phosphate concentrations. Diurnal variation in serum phosphate occurs with concentrations lowest at 9AM and highest at 7PM. There is also a seasonal variation, whereby levels are higher during the summer than during the winter; this may arise because phosphorus absorption is stimulated by the greater synthesis of vitamin-D with higher summer sunlight exposure. Serum phosphate is also higher in women than in men (by about 0.31mg/dL, 0.1mmol/L), and higher in children (i.e., normal range 1.45-1.78mmol/L) than adults (i.e., normal range 0.87-1.45mmol/L); see Table 23b.1.

Normal serum phosphate values for children and adults are shown in Table 23b.1. Hypophosphatemia is usually defined as serum phosphate < 0.87mmol/L and hyperphosphatemia as a serum phosphate > 1.45mmol/L; see Section 23b.8 for more discussion of abnormalities in serum phosphate (Koumakis et al., 2021).

This page titled 24.3: Interpretive criterion- serum phosphate (23b.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.4: Phosphate balance (23b.4)

Phosphate balance is the result of the interaction of intestinal absorption of dietary phosphate, renal phosphate excretion, and exchange of phosphate between extracellular and bone and intracellular phosphate pools (Figure 23b-2.). At present, the only easily available parameter to study total body phosphate in a clinical setting is to measure the serum phosphate concentration, usually in the fasting state. This measurement represents only a small portion of the total body phosphate, as noted above, and can also be affected by shifts of phosphate between intracellular and extracellular compartments (Uribarri & Oh, 2018).

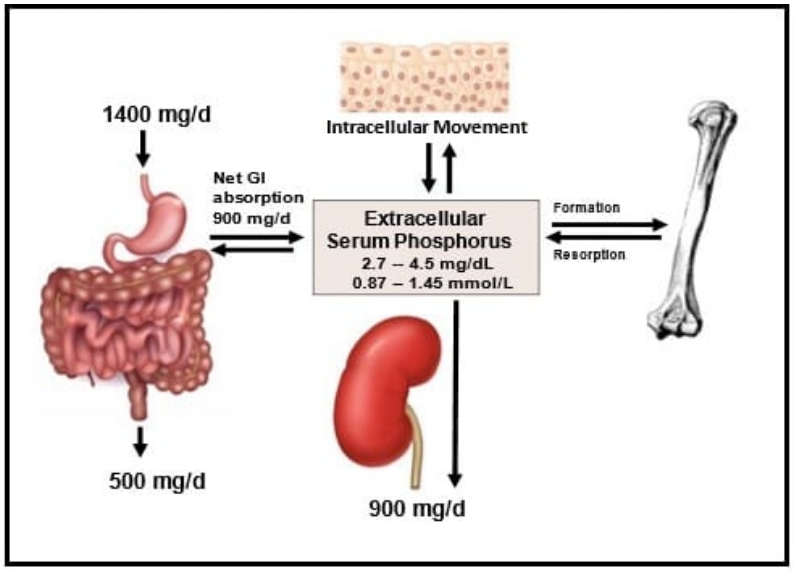


Figure 23b.2 The diagram illustrates phosphorus balance maintained in a healthy adult consuming an average American diet containing 1400mg phosphate (Pi). The components of phosphorus or phosphate balance, include intestinal absorption, kidney excretion, bone formation/resorption, the intracellular space, and plasma transport. Under conditions of normal renal function, the amount of Pi absorbed equals the amount excreted in the urine, thus balance is achieved in an adult where bone formation equals bone resorption and tissue uptake equals that released.

The kidneys play a major role in phosphate balance (Figure 23b.2) by adjusting urinary excretion (output) to match net gastrointestinal absorption (input) of phosphate to maintain zero balance in an adult or, to retain phosphate to maintain positive balance in a child for growth, or for a pregnancy. In healthy subjects, the kidneys reabsorb about 89% of the filtered load of phosphate, with the rest being excreted in the urine. Plasma phosphate filtered in the glomerulus is mainly reabsorbed in the proximal renal tubules (75%), with only 10% reabsorbed in the distal tubules, leaving about 10-15% in the urine.

Gastrointestinal (GI) phosphate absorption in humans has traditionally been measured as the difference between dietary and fecal phosphate content; this net phosphate absorption is a linear function of dietary phosphate intake(IOM, 1997). For a dietary phosphate intake within the range of 4-30mg/kg/day (280-2100mg per day for an adult), net absorption is about 60-65%. Shown in Figure23b.3 are the two main transport systems for intestinal phosphate absorption: one is an active, sodium-dependent, saturable and transcellular transporter, and the other is a passive, sodium-independent, non-saturable and paracellular transporter (Marks, 2019). The intestinal sodium-dependent transporter is regulated by vitamin D and parathyroid hormone (PTH) and is often referred to as “active” transport. In contrast, the paracellular phosphorus absorption pathway lacks a tight regulation and depends on the phosphate concentration gradient across the epithelium, the electrical gradient (lumen negative), and tight junction permeability (Calvo & Uribarri, 2021).

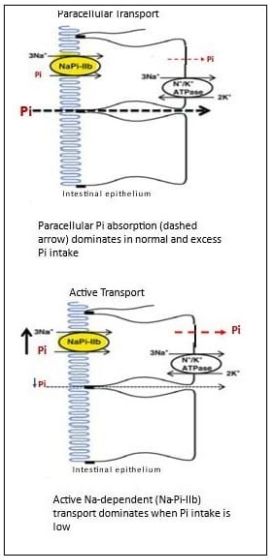


Figure 23b.3 Mechanisms of Intestinal phosphate Absorption. Modified from Marks (2019).

Most organic and inorganic phosphate is absorbed in the small intestine after liberation by gut and intestinal enzymes. However, the dominant plant source of phosphate, phytate (Figure 23b-1), is poorly absorbed in humans because of lack of the enzyme phytase. Many colonic bacteria produce phytases (myo-inositol hexakisphosphate phosphohydrolases) capable of sequentially hydrolyzing phytate, releasing phosphate. Liberated inorganic phosphate then has the potential to be absorbed in the colon via paracellular transport, although the overall importance of this remains uncertain. Increasing solidity of distal colon fecal contents could potentially make soluble phosphate less accessible for paracellular absorption. Nevertheless, a recent review concluded that at least 50% of phosphate present in phytate is recovered as phosphate in 24-hour urine collections based on results of a series of earlier human studies (Calvo & Uribarri, 2021).

This page titled 24.4: Phosphate balance (23b.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.5: Hormonal regulation of serum phosphate (23b.5)

Serum phosphate concentration must be maintained within a very narrow range to avoid adverse health consequences and risk of disease such as soft tissue calcification or cardiovascular disease. Regulation of serum phosphate involves the interplay of four organs, phosphate membrane transporters bound in these organs, and the actions of three endocrine hormones that influence the activity of the phosphate transporters(Uribarri & Calvo, 2023). The four major organs involved in regulating serum phosphate are the kidneys, bone, intestine, and parathyroid glands. There are two families of sodium-phosphate membrane transporters specific to these organs. They include the SLC34 group of NaPi-2a, b, c, chiefly located in the kidney and intestine, and the SLC20 family (PiT-1 and 2) largely found in bone, intestine, soft tissue, muscle, with some in the kidney (Forster et al., 2013). The activity of these two families of sodium-phosphate membrane transporters is controlled by three endocrine hormones: parathyroid hormone (PTH), calcitriol (1-25,dihydroxycholecalciferol; (1-25$(OH)_2$D), the active metabolite of vitamin D), and fibroblast growth factor 23 (FGF-23). The three phosphate regulating hormones are endocrine hormones meaning that they are secreted into the circulation by a specific organ, but act upon a distal organ.

As illustrated in Figure 23b.4, PTH is secreted by the parathyroid glands when the rise in serum phosphate triggers a decrease in serum ionized calcium or is sufficiently elevated to directly stimulate PTH secretion. Circulating PTH rapidly acts to decrease NaPi-2a and NaPi-2c co-transporters in the renal proximal and distal tubules. A decrease in membrane co-transporters acutely decreases phosphate reabsorption and increases phosphate excretion in the urine. When normal serum phosphate concentration is filtered in the glomerulus, about 75% of phosphate is reabsorbed in the proximal tubule and 10% from the distal tubule with 10-15% lost in urine, as noted earlier.

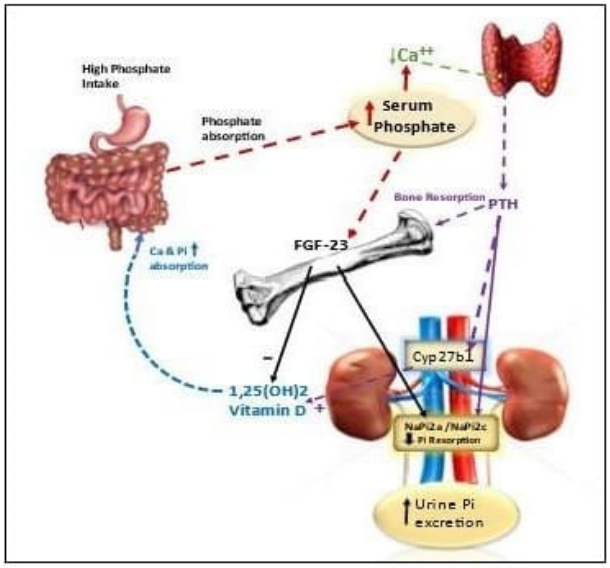


Figure 23b.4 Stimulatory pathways in the diagram are shown by dashed lines and inhibitory pathways by solid lines in red for hyperphosphatemia, purple for PTH, black for FGF-23 and blue for 1,25-dihydroxy vitamin D (calcitriol).

The action of PTH rapidly lowers serum phosphate by increasing urinary loss, and rapidly increases serum calcium and to a lesser degree phosphate by stimulating bone resorption. PTH action more slowly restores serum calcium through up-regulation of the renal cytochrome enzyme *Cyp27b1* that catalyzes the activation of 25-hydroxy vitamin D to the active metabolite, calcitriol (1,25$(OH)_2$D) secreted by the kidney into the circulation. In turn, the circulating hormonal form of vitamin D, calcitriol, acts on the small intestine to increase active calcium transport, thus PTH rapidly corrects serum phosphate and calcium concentrations that

stray from the normal range (Uribarri & Calvo, 2023).

When excessive intake of phosphate is sustained over time, or kidney function fails resulting in hyperphosphatemia, hormonal control of serum phosphate is reliant on regulating intestinal phosphate absorption and renal tubular phosphate reabsorption and may require the action of the bone secreted hormone FGF-23 (Raush & Foeller, 2022). In response to hyperphosphatemia, FGF-23 is secreted by osteocytes in bone and similar to PTH, FGF-23 acts to suppress co-transporter action (proximal tubule NaPi-2a and NaPi-2c), decreasing renal reabsorption and increasing urinary phosphate excretion but acts to supress PTH secretion (not shown) . In contrast to the action of PTH, FGF-23 inhibits 1,25-dihydroxy vitamin D renal synthesis and intestinal phosphate absorption by downregulating renal *cytochrome P450 (Cyp27b1)* expression (the key enzyme for calcitriol production), and enhancing renal *Cyp24a1* production, thus catalyzing the inactivation of 1,25-dihydroxy vitamin D.

FGF-23 action may be dependent or independent of Klotho, a beta-glucuronidase enzyme that occurs as both a transmembrane protein and a secreted renal protein, and which can function as a co-receptor to FGF-23 (not shown in Figure 23b.4). Although many of Klotho's functions remain unclear, it has a proven role in phosphate regulation (Kuro-O, 2019). FGF-23 clearly suppresses calcitriol synthesis and, with time, PTH secretion; however, initially in response to high serum phosphate, PTH is believed to stimulate FGF-23 secretion from osteocytes (Raush & Foeller, 2022). Not shown in the simplified diagram of hormonal control of serum phosphate (Figure 23b.4), calcitriol upregulates Na-Pi co-transporters (an action opposite to that of FGF-23) in both the intestine and proximal renal tubules which leads to both increased intestinal absorption of phosphate and increased renal phosphate reabsorption. Calcitriol suppresses PTH secretion and has a negative feedback action on renal 1-alpha hydroxylase *(Cyp27b1)*, thereby reducing its own production. With lower calcitriol concentrations, 24-hydrolxylase is upregulated, increasing the production of 24,25-dihydroxy vitamin D effectively inactivating calcitriol hormonal action.

This page titled 24.5: Hormonal regulation of serum phosphate (23b.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.6: Nutrient reference values (23b.6)

In 1997, the Institute of Medicine (U.S. and Canada) determined adult age specific Estimated Average Requirement (EAR) values for phosphorus derived from studies using serum phosphorus as a biomarker (IOM, 1997). For adults 19-50 years, the EAR was based on the relationship between serum phosphate and absorbed intake determined from earlier published data. This relationship was then used to translate absorbed intake to the amount of ingested phosphorus based on an assumed efficiency of absorption of 62.5% from a mixed diet not high in phytate (Calvo & Whiting, 2018). An intake of phosphorus of 580mg/d meets the needs of 50% of the adult population (≥ 19y) and therefore was set as the EAR and served as the basis for determining the Recommended Daily Allowance (RDA) for phosphate that covers the phosphorus needs of 97% of the adult population. For children, the biomarker used to set the EAR for dietary phosphorus intake was based on published factorial estimates of accretion of phosphate into bone; see IOM (1997) for more details. For infants, the AI was set to reflect the observed mean intakes of infants fed principally with human milk (IOM, 1997).

The EAR and RDA values for phosphorus for infants, children and adults by age and sex, and for the physiologic states of pregnancy and lactation recommended in 1997, are shown in Table 23b.2.

*Table 23b.2 US and Canadian phosphorus Dietary Reference Intakes (mg P/day). Estimated Average Requirement (EAR), Recommended Dietary Allowance (RDA, or Adequate Intake equivalent) and Upper Level (UL) for phosphorus * Applies to pregnant and lactating women. Source: (IOM, 1997).*

Age/sex groups	EAR	RDA	UL
0-6 mo	--	100 (AI)	-
6-12 mo	--	275 (AI)	-
1-3y F & M	380	460	3000
4-8y F & M	405	500	3000
9-18y F & M	1055	1250	4000
19-50y F* & M	580	700	4000
51-70y F & M	580	700	4000
71+ F & M	580	700	3000

A tolerable upper intake level (UL) for phosphate was also set in 1997 as part of the Dietary Reference Intakes (DRIs). The UL is considered to be a safe intake level, but as intake increases above the UL, risk for adverse events increases (IOM, 1997). The UL for adults is set at 4000mg/d. Researchers now question the method used by the IOM panel that set a value nearly six times higher than the RDA for phosphorus, suggesting a need to revisit the UL (Whiting & Calvo, 2018). The increased use of phosphate-containing food additives in processing led the European Food Safety Authority (EFSA) to reassess the safety of phosphate additives and cumulative phosphate intakes. The 2019 EFSA review resulted in a revision of the “group acceptable daily intake” (ADI) of 4.2g phosphate per day to the lower value of 2.8g phosphate per day for an average 70kg adult (EFSA, 2019).

Current mean dietary phosphate intakes of Americans are 2-3-fold higher than the RDA for all age groups > 1y (with the single exception of rapidly growing adolescents), but do not exceed the current UL. Nonetheless, excess phosphate intake is associated with growing evidence of potential noncommunicable disease risk. This is of concern especially when considering the high intakes of bioavailable phosphate that may occur from the consumption of ultra-processed foods. Consequently, consideration should be given to revising the 1997 DRI’s (notably the UL) (Uribarri & Calvo, 2023).

This page titled 24.6: Nutrient reference values (23b.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.7: Dietary sources

Dietary sources of phosphate include both organic and inorganic phosphate. Phosphate is present in most food sources and is usually highest in animal protein compared to plant protein sources. Importantly, the bioavailability of phosphorus differs between protein derived from animal sources compared to plant derived protein. In addition to containing protein, plants store phosphate as phytate (Figure 23b.1). Phosphate is stored as phytate in unrefined cereals, oil seeds, and legumes that requires enzymatic or physical action to release phosphate bound to phytate. Bioavailability of phytate-phosphorus is low unless foods are processed using home-based methods such as soaking, germination, or fermentation, or commercial methods such as canning and extrusion (Tefarra, 2021), all of which having the potential to improve phosphate availability from phytate. While these methods of processing improve phosphate availability from phytate, commercial processing leading to processed and ultra-processed foods may involve the addition of phosphate-containing additives that provide highly available sources of phosphate, such as sodium phosphate salts shown in Figure 23b-1 (Calvo & Uribarri, 2021).

Examples of animal, plant, and additive sources of phosphate are shown in Table 23b.3, with the foods classified according to the NOVA system (Monteiro et al., 2017), which reflects their degree of processing. The NOVA system classifies foods into four groups, with the fourth group called “ultra-processed”. A recent cross-sectional study of the United States food supply found that 71% of packaged foods and beverages were ultra-processed and 60% of energy intake came from ultra-processed foods, an increase over past decades (Baldridge et al., 2019).

High ultra-processed food consumption patterns are associated with increased risk of cardiovascular disease and mortality (Juul et al., 2021), chronic kidney disease (Cai et al., 2022), and various other serious chronic diseases. Specifically, ultra-processed foods such as cured meats containing added inorganic phosphates are associated with carotid intimal thickening (Itkonen et al., 2014), decreased renal function (Duong et al., 2022), and low HDL cholesterol (Fulgoni et al., 2022). Causality between phosphate additive use in ultra-processed foods and increased risk of chronic disease or mortality remains to be established. However, there is growing awareness that the use of phosphate additives in food processing increases the total phosphate availability and content of foods. As can be seen in Table 23b.3, it is difficult to describe dietary sources of phosphate as high or low without information on the source and form of phosphate, and their relative bioavailability. In general, both plant and animal protein sources are good sources as noted earlier. In addition, the phytate-phosphorus present mainly in unrefined cereals, oil seeds, and legumes may become bioavailable, if during processing phytate has been hydrolyzed, and phosphate bound to phytate released prior to absorption. Under the Nova classification, raw and minimally processed phytate rich plant foods have lower phosphorus bioavailability than comparably processed animal protein sources (Calvo & Uribarri, 2021).

*Table 23b.3 Characteristics and sources of phosphate (organic (OPi), inorganic (IPi) or phytate) in foods classified according to degr foods. ** Indicates frequent use > 1 IPi additive + processed protein isolates or concentrate. No determination (ND) made for Group*

NOVA Food Classification Definitions	Representative Foods (Prepared phosphate Conte
Group 1. Minimally Processed, Unprocessed Foods: Natural plant and animal foods and water processed by drying, crushing, grinding, fracturing, filtering, roasting, boiling non-alcoholic fermentation, pasteurizing, refrigeration, chilling, freezing, placing in container and vacuum packaged for storage	Cooked oatmeal (99) Boiled lentils (180) Boiled black beans (140) Soy flour (674) Corn flakes breakfast-cereal (58)
	Baked chicken breasts (228) Baked tuna fillet (326) 2% Milk (92) Broiled beef steak (189) Boiled egg (126)
Group 2. Processed Culinary Ingredients Includes oils, butter, sugar, and salt that allows for preparing unprocessed foods at home	Soy oil (0) Butter (24) Sugar (0) Salt (0)

NOVA Food Classification Definitions	Representative Foods (Prepared phosphate Conte
Group 3. Processed Foods: Foods processed by various preservation or cooking methods or fermentation such as bread that usually contains 2 to 3 ingredients.	Grilled Chicken Patty, from frozen (208) Cheddar Cheese (473) Canned Tuna* (311) Canned Green Beans* (19)
Group 4. Ultra-processed Foods: Foods industrially formulated from substances derived from foods, additives, and ultra-processed products not usually used in home cooking such as hydrogenated oils, hydrolyzed proteins, added sugars and salt	Fried chicken nuggets*(213) Fried Sausages** (174) Battered fish sticks*(171) Baked frozen pizza* (215) Processed Cheese* (982) Cheerios breakfast-cereal** (448) Pancakes from Mix* (334) Shop white bread** (109)

In minimally processed and unprocessed foods (Group 1, figure 23b.3), the estimated percent phosphate absorption varies from less than 40% to 70% being generally lower in plant foods and higher in animal foods. There is very little data on the estimated phosphate absorption from processed culinary ingredients, but processed foods (Group 3, figure 23b.3) consumed as part of a mixed diet, can have an estimated percent phosphate absorption of greater than 60%. The absorptionn in ultra-processed (Group 4, figure 23b.3), animal foods may range from 90-100% but is lower in ultra-processed plant foods.

The presence of phosphate additives in minimally and ultraprocessed foods (shown in Table 23b.3 by *) may or may not show differences in total phosphate content compared to unprocessed foods if the content estimates are based on national nutrient content database information (Calvo, et al., 2014). There is a lack of accounting for the phosphorus contribution by phosphate additives in most foods in the nutrient content databases resulting in a widely recognized underestimation of phosphorus intake (Calvo, et al., 2019). Regrettably, many investigators have only recently discovered the need to directly analyze the phosphate content of the foods fed in studies to accurately design clinical trials examining dietary phosphate metabolism (Stremke et al., 2018).

The phosphorus content of packaged foods in the US and Canada does not require phosphorus content to be listed on the label; however, if phosphate additives were added, then their use must be identified but not quantified in the label ingredients list (Calvo et al., 2019). The US Food and Drug Administration (FDA) requires that all ingredients (additives) added to food during processing must have prior approval for use in one of the 32 specific technical functions recognized by the FDA. Phosphate-containing additives and processed protein ingredients (isolates and hydrolysates) are approved for at least 26 of these technical functions. Processed foods often contain one or more phosphate additives (Sullivan et al., 2007), because the more frequently used additives are approved for more than one technical function. For example, of the more than 60 frequently used phosphate additives more than 30 are approved for use as a nutrient supplement, 24 for use as a stabilizer or thickener, 20 as emulsifiers and emulsifier salts and 18 as pH control agents. Understanding the role of dietary phosphate intake in chronic disease risk requires better understanding of the role of phosphate additives in food processing as phosphate additives can be added to foods for multiple functions, each contributing to higher phosphate content which is not always captured by the basic tools used by nutritional scientists to determine total phosphate intake. A simple solution to this inaccuracy in the nutrient content databases is to require phosphorus content on the Nutrient Facts Label (Calvo et al., 2019).

This page titled 24.7: Dietary sources is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.8: Abnormalities in serum phosphate (23b.8)

Phosphate balance, a proxy for assessing phosphate nutritional status, is best assessed by measuring fasting serum phosphate concentrations. Nevertheless, single fasting serum phosphate measures do not always reflect the body phosphate stores because acute shifts of phosphate between body compartments may temporarily affect serum phosphate without affecting total body phosphorus content.

Table 23b.4 Mechanisms Causing Hypophosphatemia Data source: Koumakis et al., (2021).

Paths to Acute Acquired Hypo-phosphatemia	Causative Mechanism
Acute Shift in Extra-cellular to Intra-cellular Distribution	Refeeding syndrome Metabolic acidosis Glucose infusion and other carbohydrates Acute respiratory alkalosis Alcohol withdrawal Serious burns, surgical trauma Hyperthermia Hormonal or other agents : (insulin, glucagon, cortisol, catecholamines, fructose) Rapid cell proliferation or uptake (hungry bone syndrome, cancer)
Decreased Dietary Intake	Severe dietary restriction and malnutrition (renal failure) Anorexia Kwashiorkor (severe protein / calorie malnutrition)
Decreased Intestinal Absorption	Vitamin D deficiency Antacid overuse Phosphate binder use Gastrointestinal malabsorption
Increased Renal Excretion	Primary hyperparathyroidism Secondary hyperparathyroidism (dietary phosphate excess and low calcium intake or vitamin D deficiency) Metabolic acidosis (volume expansion, Fancon syndrome, tumor production of PTH-related peptide, neurofibromatosis; acute falciparum malaria, and various medications including bisphosphonates such as etidronate, pamidronate, zoledronic acid for post-menopausal osteoporosis)

Clinically, patients may present with either hypophosphatemia (serum phosphate < 0.87mmol/L), or hyperphosphatemia (serum phosphate > 1.45mmol/L), with each one of these two conditions characterized by different manifestations and causes that cannot always be attributed to issues with intakes of dietary phosphate. The differences in clinical manifestations and causes for hypophosphatemia and hyperphosphatemia are summarized in Tables 23b.4 and 23b.5 (Koumakis et al., 2021).

Clinical manifestations of hypophosphatemia include muscle weakness, cardiomyopathy, respiratory insufficiency, osteomalacia, rickets, blood disorders, nervous system dysfunction, hypercalciuria and impaired insulin secretion. Acute hyperphosphatemia is associated with hypocalcemia, metastatic calcification, the gradual progression of renal failure, and secondary hyperparathyroidism (Uribarri & Calvo, 2023).

Transcellular phosphate shifts such as those occurring with acute alkalosis can produce significant hypophosphatemia by the intracellular shift of phosphate, even though the total body phosphate content is unaffected. These shifts are usually transient in nature; however, when hyperphosphatemia or hypophosphatemia are chronic, there is usually a correlation between serum phos-

phate levels and total body phosphate. Moreover, there is a significant circadian variation in serum phosphate concentration, as well as gender and age differences in serum phosphate unrelated to phosphate intake effect on balance. Given these problems with significant variation in serum concentration, the European Food Safety Authority (EFSA, 2015). as well as other countries did not consider serum phosphate concentration to be an appropriate biomarker for establishing dietary phosphate requirements or nutritional status.

Table 23b.5 Causes of Hyperphosphatemia. Data source: Koumakis et al. (2021)

Duration of Hyperphosphatemia	Mechanism Associated with Hyperphos
Acute Hyperphosphatemia	Acute kidney injury Increased intestinal absorption of excess dietary load phosphate-containing enemas Transcellular shifts from intra- to extracellular (hemolysis, rhabdomyolysis, acidosis, tumorlysis) Hyperthermia Heat stroke
Chronic Hyperphosphatemia	Chronic kidney disease Hypoparathyroidism Pseudohypoparathyroidism Vitamin D intoxication Disorders of magnesium regulation (PTH regulation) Acromegaly

In contrast, the Institute of Medicine of North America adopted the measurement of serum inorganic phosphate as an acceptable and easy to monitor indictor for determining phosphate requirements for adults in the U.S. and Canada (IOM, 1997).

This page titled 24.8: Abnormalities in serum phosphate (23b.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

24.9: Inherited and tumor-induced disorders of phosphate metabolism

Genetic-related disorders of phosphate metabolism , notably chronic hypophosphatemia, are very rare in most populations, Moreover, they are unrelated to dietary phosphate intakes, unlike the disorders associated with excessive phosphate intakes. These genetic phosphate disorders include X-Linked Hypophosphatemic Rickets (XLH); Autosomal-dominant Hypophosphatemic Rickets (ADHR); Autosomal Recessive Hypophosphatemic Rickets (ARHR), and Hereditary Hypophosphatemic Rickets with Hypercalciuria (HHRH). They are inherited diseases or result from missense mutations which all impact FGF-23 metabolism, by either increasing circulating levels of the hormone or its activity, which ultimately lead to hypophosphatemia arising from phosphate wasting (i.e., excessive urinary phosphate excretion in relation to serum phosphate level). Sustained hypophosphatemia can manifest as rickets in children and osteomalacia in adults but can be treated with oral phosphate supplementation, and more recently with biologics that block FGF-23 renal hormone action (Athonvarangkul & Insogna, 2021).

Tumor-induced osteomalacia (also known as oncogenic osteomalacia) is also a disorder of phosphate wasting, but it is not inherited, instead stemming from excessive levels of FGF-23 secreted by tumors. Such excessive levels of FGF-23 again promote renal phosphate wasting resulting in under-mineralized bone (osteomalacia), bone pain, fractures, and muscle weakness in adults (Brandi et al., 2021).

This page titled 24.9: Inherited and tumor-induced disorders of phosphate metabolism is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25: PHOSPHORUS (CHAPTER 23B)

Abstract

Magnesium is a cofactor for over 600 enzymatic reactions vital for life and is a controlling factor in nerve transmission, skeletal and smooth muscle contraction, cardiac excitability, vasomotor tone, blood pressure, and bone turnover. Thus, magnesium deficiency can have detrimental consequences that include impaired physical and mental well-being and risk for chronic disease. Before 1990, a nutritional magnesium deficiency was considered rare and to be present mainly in clinical conditions associated with gastrointestinal or renal loss of magnesium. However, numerous reports of an association between a low magnesium intake and chronic diseases, especially those associated with chronic inflammatory stress, has led to the recognition that mild or subclinical magnesium deficiency, also known as chronic latent magnesium (CLMD) deficiency, may be quite prevalent. As a result, magnesium has become a nutrient of public health concern and a simple, rapid, and reliable clinical measure is needed to assess for magnesium deficiency. Numerous methods of magnesium status assessment have been developed, but all have shortcomings that impact their use in the clinical setting to accurately assess magnesium status.

The predominant method for assessing magnesium status at present is the determination of total serum or plasma magnesium. However, this method is not a sensitive indicator of body stores of magnesium. In the reference range of serum or plasma magnesium values, there is an interval in which individuals may have either CLMD or an adequate status. Combining measurement of total serum or plasma magnesium with both the determination of magnesium from dietary intake and urinary excretion appears to enhance the usefulness of this method to assess magnesium status.

Recent models of ion-selective electrodes have enabled the measurement of plasma ionized magnesium, the physiologically active form, to be acceptably accurate and precise using whole blood. As a result, this method is becoming more prevalent in the clinical setting because the procedure is simpler and requires less blood than the measurement of total magnesium in plasma or serum magnesium. However, plasma ionized magnesium has shortcomings similar to those for total plasma and serum magnesium regarding the sensitivity of the reference range to determine CLMD accurately.

Measurements of magnesium in erythrocytes, erythrocyte membranes, and mononuclear cells give a reasonable assessment of physiologically active body stores of magnesium. However, all these methods are laborious and prone to error, and hence only suitable for use in a research setting and not for routine clinical use.

Urinary magnesium excretion is an excellent indicator of magnesium intake. However, a single urinary magnesium determination may not reflect magnesium stores because it responds so rapidly to changes in dietary intake. Hence, it is mainly appropriate for use in population studies, or in combination with other measures of status.

The magnesium load test provides the most valid assessment of magnesium status. It has been used to identify magnesium deficiency in elderly individuals, in chronic alcoholism, and in several chronic diseases in which hypomagnesemia was not present. The test determines the percentage of urinary magnesium retained over a given period of time after parenteral administration of a magnesium load and yields an abbreviated balance determination. The test is invasive, time-consuming, and cumbersome; it requires close supervision for at least 24 hours after a magnesium load. Because of these drawbacks, it has been used mostly as a research tool.

Several other methods used to assess magnesium status also have drawbacks that preclude their use as routine clinical methods. Muscle magnesium determination is extremely invasive and requires skill to obtain suitable samples for analysis. Buccal cell magnesium determinations also require skill in obtaining cells, and special expensive equipment for analysis. Magnesium balance requires controlled and consistent dietary intakes and careful collection of urine and stool over a lengthy study period. Fractional excretion of magnesium requires the determination of three variables and has no apparent advantage over other well-established methods. The magnesium depletion score needs further evaluation and validation, although seems to be most useful for establishing status in individuals with a chronic disease. However, all these methods do not have an established validated reference range that indicates magnesium adequacy or deficiency.

At present, there is no simple, rapid, and reliable single clinical method to determine the presence of chronic, latent magnesium deficiency (CLMD). Determination of serum total or ionized magnesium remains the most acceptable choice in the clinical setting. The reliability of these measures of status could be improved if their values for the reference ranges indicative of deficient or adequate magnesium status are confirmed by other measures such as the determination of magnesium in both dietary intakes and urinary excretion.

25.1: Introduction (23c.1)
25.2: Functions of magnesium (23c.2)
25.3: Absorption and metabolism (23c.3)
25.4: Magnesium deficiency in humans
25.5: Food sources and dietary Intakes
25.6: Effects of high magnesium intakes (23c.6)
25.7: Occurrence of dietary deficiency (23c.7)
25.8: Indices of magnesium status (23c.8)

This page titled 25: Phosphorus (Chapter 23b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.1: Introduction (23c.1)

The human body contains about 760mg of magnesium at birth, 5g at age 4-5 months, and 25g when adult (World Health Organization and Food and Agriculture Organization of the United Nations, 2004). There are three body pools of magnesium (Rude & Shils, 2006). One pool has a turnover rate of less than 28hours. This pool of extracellular magnesium, which includes blood, contains 0.8-1.0% of body magnesium. The second pool has a turnover rate of about 11days. This pool is primarily intracellular magnesium, of which about 40% of the magnesium is found in soft tissues where it performs most of its essential functions. Skeletal magnesium is the third pool with a slow turnover rate measured in months or years in normal adults. Skeletal magnesium contains about 60% of body magnesium of which two-thirds is found within the hydration shell (bone mineral or poorly crystallized hydroxyapatite) and one-third on the crystal surface of the skeleton (Rude & Shils, 2006). Surface magnesium is readily exchangeable with serum and serves as a reservoir for the maintenance of magnesium for essential functions when intakes of magnesium are deficient.

This page titled 25.1: Introduction (23c.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.2: Functions of magnesium (23c.2)

Divalent magnesium ($Mg2^{+}$) is the fourth most abundant cation in the body and is second to potassium as the most abundant intracellular cation. Magnesium is a cofactor for over 600 enzymatic reactions vital to pathways including DNA, RNA, protein and adenosine 5'-triphosphate (ATP) synthesis; cellular energy production and storage; glycolysis; and cellular second messenger systems. The predominant enzymatic role of magnesium in these pathways involves an enzyme interacting with MgATP to complete a reaction. Magnesium also binds directly to enzymes to cause conformational change (allosteric activation) that gives a catalysis center for reaction to occur. Stabilization of ribonucleotides and deoxyribonucleotides by magnesium allow for DNA duplication, transcription, and maintenance, and transfer RNA function (Rude & Shils, 2006). Magnesium reacts with phosphates and carboxylates to stabilize membranes, affecting their fluidity and permeability which influences cellular ion channels, transporters, and signaling (Rude & Shils, 2006). Through regulating $Ca2^{+}$ and K^{+} movement in and out of the cell, magnesium is a controlling factor in nerve transmission, skeletal and smooth muscle contraction, cardiac excitability, vasomotor tone, blood pressure, and bone turnover.

This page titled 25.2: Functions of magnesium (23c.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.3: Absorption and metabolism (23c.3)

The jejunum and ileum are the primary sites of magnesium absorption; lesser amounts are absorbed by the cecum and colon. When dietary intakes of magnesium are near requirement, 45-55% is absorbed. With intakes less than 50% of requirement, the percentage of magnesium absorbed may increase to 65-70%. As dietary intake increases over requirement the percentage of magnesium absorbed decreases. For example, with an intake of 972mg, fractional absorption of magnesium was only 11% (Fine et al., 1991).

Magnesium is absorbed via both active transport (transcellular) and passive diffusion (paracellular). Active transport occurs in the ileum, although apparently also in the cecum and colon through the action of magnesium transporters: transient receptor potential melastatin (TRPM) type 6 (TRPM6) and TRPM7 (de Baaij et al., 2015). Passive paracellular diffusion accounts for about 90% magnesium absorption and is dependent on a high concentration in the jejunum and ileum and a transcellular potential difference generated by sodium transport (Romani, 2013).

Magnesium absorption can also be affected by other dietary components, including calcium, phosphorus, fiber, phytate, protein, and oxalates, all of which can decrease absorption, whereas vitamin D and its active metabolites, carbohydrates fermentable by bacteria, and oligosaccharides increase absorption (Nielsen, 2017).

The primary organ regulating magnesium homeostasis is the kidney where about 10% of total body magnesium is filtered daily through the glomeruli with the nephron recovering 95-99% of filtered magnesium. Passive diffusion associated with calcium, sodium, and water transport results in the reabsorption of about 10-25% of the magnesium (de Baaij et al., 2015). Passive diffusion in the thick ascending limb of the loop of Henle accounts for about 50-70% of the magnesium reabsorption. The remaining 5-10% of filtered magnesium is reabsorbed in the distal convoluted tubule via an active transcellular mechanism. The active transport channel in the kidney apparently involves TRPM6. When magnesium intakes are deficient, excretion of magnesium by the kidney decreases to 12-24mg/day to conserve body magnesium. Diets high in sodium, calcium and protein, and caffeine and alcohol consumption all enhance magnesium excretion (Costello & Rosanoff, 2020).

This page titled 25.3: Absorption and metabolism (23c.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.4: Magnesium deficiency in humans

Signs and symptoms of severe magnesium deficiency include positive Trousseau's and Chvostek's signs; muscle spasms, fasciculations, and tremor; and personality changes such as apathy, depression, nervousness, delirium, and hallucinations (Rude, 1998). However, severe magnesium deficiency involving these signs and symptoms is rare and usually associated with predisposing and complicating disease states that cause low magnesium intakes or impaired intestinal or renal absorption.

In conditions of long term moderate or mildly deficient intakes of magnesium or excessive excretion, magnesium is slowly released from bone to assure that vital essential functions are fulfilled. Nevertheless, there may still be a lack of magnesium required for some functions, but the signs associated with severe magnesium deficiency outlined above do not occur. This lack has been called chronic latent magnesium deficiency (CLMD)(Elin, 2010).

Chronic latent magnesium deficiency is considered a factor contributing to several chronic diseases, especially cardiovascular disease, including ischemic heart disease, sudden cardiac death, heart failure, atrial fibrillation, and ischemic stroke (Costello & Rosanoff, 2020). The associated risk of CLMD with cardiovascular disease is through impaired glucose metabolism and insulin action, dyslipidemia, hypertension, chronic inflammation, and impaired vasomotor tone (Costello & Rosanoff, 2020).

Chronic latent magnesium deficiency also may result in an increased risk for bone fractures and osteoporosis (Costello & Rosanoff, 2020). Apatite crystals formed in magnesium-deficient bone are increased in size, which makes bone more brittle, increasing the fracture risk. Magnesium deficiency can also reduce parathyroid hormone and 1,25-dihydroxycholecalciferol levels, which affect the regulation of magnesium-dependent calcium channels and changes in the release of inflammatory cytokines. Such effects enhance osteoclastic action and inhibit osteoblastic action, resulting in bone loss and an increased risk for osteoporosis.

Other chronic diseases with which CLMD has been associated include osteoarthritis, chondrocalcinosis, depression, and migraine headaches(Costello & Rosanoff, 2020). In addition, CLMD has been found to cause chronic inflammatory & oxidative stress, indicating that through these effects, CLMD may contribute to chronic diseases, other than those noted above, in which inflammatory or oxidative stress facilitates their occurrence.

This page titled 25.4: Magnesium deficiency in humans is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.5: Food sources and dietary Intakes

Foods of plant origin provide about 50% of the magnesium intake for adults on mixed diets (Hunt & Meacham, 2001). Meats supply about 14-16% and dairy products (including milk) supply about 34% of the magnesium intakes in U.S. adults (Hunt & Meacham, 2001). Rich food sources of magnesium include whole grains / cereal, nuts, pulses, green leafy vegetables, and dark chocolate. Intermediate sources of magnesium include cheese, meats, and most seafoods. Refined grains are poor sources of magnesium.

This page titled 25.5: Food sources and dietary Intakes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.6: Effects of high magnesium intakes (23c.6)

Because the kidneys are very effective in the excretion of excessive magnesium, severe magnesium toxicity rarely occurs. When severe toxicity resulting in high blood magnesium occurs, the condition arises almost exclusively in individuals with kidney dysfunction or failure who are given magnesium salts or drugs such as laxatives or antacids. Symptoms of toxicity include lethargy, confusion, nausea, diarrhea, impaired breathing, hypotension, muscle weakness, and heart arrhythmias (Romani, 2013). Signs of hypermagnesemia also include low blood calcium and high blood potassium. The heart rhythm changes in magnesium toxicity are caused by abnormal electrical conduction which occurs at the nervous, muscular, and cardiac level.

Severe magnesium toxicity caused by excessive intake through food has not been reported. High intakes of supplemental magnesium available in forms such as aspartate, citrate, chloride, gluconate, lactate, and oxide can cause adverse effects such as diarrhea, abdominal cramping, and nausea. A tolerable upper intake level (UL), defined as the highest level of daily intake that is likely to pose no risks of adverse health effects to almost all individuals in the general population, was set at 350mg of supplemental magnesium in the United States and Canada for children older than 8 years and adults by the (Institute of Medicine. 1997),

This page titled 25.6: Effects of high magnesium intakes (23c.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.7: Occurrence of dietary deficiency (23c.7)

In the official Dietary Reference Intakes (DRIs) established for the United States and Canada (Institute of Medicine. 1997), the Estimated Average Requirements, (EAR, defined as average daily intake estimated to meet the daily requirements of 50% of healthy individuals) were set for individuals from aged 19 to 70 years and over at 330 to 350mg/day for males and at 255 to 265mg/day for females. These EAR values for adults are consistent with those set for adults in several other countries, as shown in Table 23c.1.

Table 23c.1 Comparison of dietary requirements of magnesium for adults (mg/day) set by four different authorities. A panel from the European Food Safety Authority (2015). considered that Average Requirements and Population Reference Intakes for magnesium cannot be derived for adults, infants or children, and therefore defined Adequate Intakes (AIs), based on observed intakes in healthy populations in the European Union (EU). EAR, Estimated Average Requirement; RDA, Recommended Dietary Allowance; RNI, Reference Nutrient Intake; RDI, Recommended Dietary Intake; RNI, Recommended Nutrient Intake; NA, Not Available.*

US/Canada Institute of Medicine (1997).	Age(y)	EAR	RDA
Males	19-30	330	400
	31-70+	350	420
Females	19-30	255	310
	31-70+	265	320
United Kingdom Dietary Reference Values (1991).	Age	EAR	RNI
Males	19-75+	250	300
Females	19-75+	200	270
Australia/New Zealand NZ Ministry of Health (2006).	Age	EAR	RDI
Males	19-30	330	400
	31-75+	350	420
Females	19-30	255	310
	31-75+	265	320
WHO (2004).	Age		RNI
Males	19-65	NA	260
	65+	NA	224
Females	19-65	NA	220
	65+	NA	190

In a United States survey performed in 2005-2006, magnesium intakes less than the EAR were reported in 48% of females in ages 31-50y, 55% in ages 51-70y, and 70% in ages 71y, and over; for males, intakes less than the EAR were 45% in ages 31-50y, 58% in ages 51-70y and 80% in ages 71y and over (Moshfegh et al., 2009).

Recently, it has been suggested that the DRIs for magnesium established for the United States and Canada should be revised and lowered to reflect findings of the improved balance study data (Nielsen, 2016). These balance study data indicated that the EAR and the Recommended Dietary Allowance (RDA, defined as average daily intake sufficient to meet the daily requirements

for almost all (about 97%) of healthy individuals) for 70kg healthy adults should be 175 and 250mg/d for 70kg healthy male and female individuals, and increase or decrease based on body weight, independent of age. The balance data found neutral balance (i.e., when intake and loss of magnesium from the body were equal) at 2.36mg/kg/day, the value indicating the amount of decrease or increase in the DRIs with each change in kg of body weight. Even with the lower DRIs, many individuals weigh more than 70kg and/or have health problems which will result in an increase in the need for magnesium. As a consequence, it is likely that over 25% of the United States population have an intake of magnesium less than the EAR. Thus, because mild or moderate magnesium deficiency can increase the risk for numerous chronic disorders, it is apparent that magnesium is a nutrient of concern requiring a clinically useful and sound indicator of magnesium status.

This page titled 25.7: Occurrence of dietary deficiency (23c.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

25.8: Indices of magnesium status (23c.8)

Because magnesium has so many critical functions, the body has strong mechanisms, such as reducing urinary excretion and releasing skeletal stores of magnesium, to assure that magnesium is readily available to fulfill those functions. In addition, mild or moderate magnesium deficiency is mostly asymptomatic. As a result of these characteristics, success in finding a cost-effective, time-efficient, and reliable indicator of magnesium status is lacking. Among the status indicators that have been evaluated are serum or plasma total magnesium, serum ionized magnesium, erythrocyte magnesium, erythrocyte membrane magnesium, blood mononuclear cell magnesium, urinary magnesium excretion, magnesium load test, sublingual magnesium, muscle magnesium, and fractional excretion of magnesium. An additional factor that may be useful in evaluating exposure to magnesium is dietary intake.

23c.8.1 Serum/plasma magnesium

At present, total serum magnesium is the predominant clinical test to assess magnesium status, although plasma total magnesium is sometimes measured instead of serum magnesium. However, serum is preferred because anti-coagulants might be contaminated with magnesium or affect assessment methods. Serum magnesium determination is inexpensive and easily performed using atomic absorption spectrophotometry or colorimetric methods. The normal range for serum magnesium is 0.75-0.95mmol/L. (Figure 23c.1). Hence, theoretically, serum concentrations above 0.75mmol/L should be considered magnesium sufficient, and those below 0.75mmol/L should be considered magnesium deficient, although, as indicated below, perhaps not rightly so (Costello et al., 2016). Serum magnesium concentrations are reportedly acceptable for determining a severe lack of magnesium intake or an elevated urinary excretion of magnesium resulting in a depletion of body stores. This finding is based on the results of a study in which exposure to extreme dietary magnesium deprivation (less than 12mg/day) resulted in a decline in serum magnesium from 0.80mmol/L to 0.61mmol/L in three weeks (Fatemi et al., 1991). However, in mild or moderate dietary magnesium deprivation, serum magnesium concentrations can remain in the normal range after several weeks, based on controlled metabolic unit studies of postmenopausal women whose dietary intakes were changed from adequacy (250mg/day) to moderate deficiency (80-160mg/day) (Nielsen & Johnson, 2017). Nevertheless, the magnesium-deprived women showed signs indicative of magnesium deficiency that responded to magnesium supplementation, indicating their body tissues were becoming magnesium deficient. Results of these metabolic unit studies are consistent with clinical studies of patients with diabetes mellitus, alcoholism, and malabsorption syndromes, all of whom had serum magnesium concentrations in the sufficient range, yet had low magnesium concentrations in erythrocytes, mononuclear cells, and muscle (Ryzen et al., 1986; Nadler et al., 1992; Abbott et al., 1994; Rude & Olerich, 1996). These findings indicate that individuals with serum/plasma magnesium concentrations within the current reference range might be deficient enough to have a bearing on the risk for chronic disease.

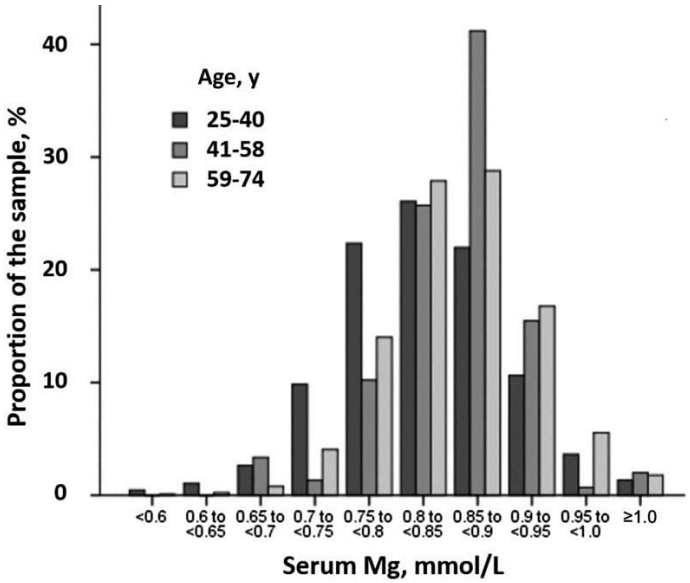


Figure 23c.1 Age-specific distributions of serum magnesium in US adults. Data were derived from NHANES I (1971 1975). Modified from (Costello et al. 2016).

Recently, it has been suggested that the reference range for serum magnesium should indicate that some individuals with serum magnesium concentrations between 0.75mmol/L and 0.85mmol/L might have CLMD(Costello et al., 2016). A review (Elin, 2010) has indicated that magnesium deficiency is present in 90% of individuals with serum magnesium concentrations of 0.70mmol/L or less, in 50% of individuals with concentrations of 0.75mmol/L, and in 10% of individuals with concentrations of 0.85mmol/L, emphasizing that serum magnesium alone lacks reliability as a status indicator of CLMD. Moreover, it also indicates that individuals with low serum magnesium concentrations might have adequate magnesium status.

The reliability of serum magnesium to indicate mild or moderate magnesium deficiency may be improved by also assessing urinary magnesium excretion and dietary magnesium intake. Metabolic unit findings suggest that individuals with serum magnesium concentrations >0.75mmol/L or as high as 0.85mmol/L could have a magnesium deficit such that they respond to magnesium supplementation if they have a dietary history showing a magnesium intake <250mg/day and a urinary excretion of magnesium of <80mg/day (Nielsen & Johnson, 2017).

Factors affecting serum magnesium

There also is a need to appreciate the numerous factors affecting serum magnesium concentrations when considering the reliability of this as a measure of magnesium status.

Diurnal variation in serum magnesium concentrations occurs: values are lower in the morning than in the evening (Wilimzig et al., 1996).

Strenuous exercise reduces serum magnesium through a shift from plasma into erythrocytes and an increase in excretion through urinary excretion and sweat loss (Lukaski, 2000).

Various drugs, such as diuretics, proton pump inhibitors, cancer therapies, and antibiotics may decrease serum magnesium concentrations (Rude, 1998), whereas antacids and cathartics may elevate concentrations (Arnaud MJ, 2008).

Disease states such as osteoporosis may decrease concentrations (Stendig-Lindberg et al., 1993). Chronic renal failure may increase concentrations by impairing magnesium homeostasis by the kidney (Elin, 1991).

Serum albumin concentrations are related to serum magnesium in a linear fashion (Kroll & Elin, 1985). Thus, hypoalbuminemia may result in low serum magnesium concentrations, and hyperalbuminemia in high serum magnesium concentrations.

23c.8.2 Serum, plasma, and blood ionized magnesium

Ionized magnesium is the physiologically active form of magnesium that is transported across cell membranes and participates in metabolic interactions and enzymatic reactions. Thus, it has received attention as a better measure than serum total magnesium for indicating the status of magnesium for essential cellular functions in the body. Ionized magnesium is determined in serum, plasma, or whole blood samples by using equipment with ion-selective electrodes. The first ion-selective electrodes were challenging to use, cost-intensive, and had poor ability to discriminate ionized magnesium from ionized calcium; this inhibited the use of ionized magnesium as a status-assessment measure. However, recent models of ion-selective electrodes are cost-effective and have been shown to be acceptably accurate and precise in measuring only ionized magnesium.

Originally, serum ionized magnesium was often determined, but at present, it is plasma magnesium that is mostly measured in whole blood; the ionized magnesium in erythrocytes is not measured. An advantage of using whole blood is that less blood and fewer procedures are needed for analyses. High pH and high concentrations of ionized calcium, lipophilic cations, and thiocyanate (caused by smoking tobacco) can interfere with the determination of plasma ionized magnesium by ion-selective electrodes (Ansu Baidoo et al., 2023).

Plasma ionized magnesium as a status assessment measure has the same problems as those outlined for plasma total magnesium; the body also attempts to maintain magnesium concentrations in plasma at a level needed for vital functions through reduced urinary excretion and bone resorption. Factors that affect plasma total magnesium, such as diuretics, proton pump inhibitors, and various disease states, also affect plasma ionized magnesium. An additional issue is that a reference range for plasma ionized magnesium indicative of adequate magnesium status has not been firmly established and appears to vary depending upon the instrument used for determining a range. A systematic review and meta-analysis involving all types of instruments used found a 95% confidence interval for ionized magnesium in healthy adults to be 0.40 to 0.68mmol/L (Ansu Baidoo et al., 2023). Using this as a reference range indicates plasma ionized magnesium concentrations below 0.40mmol/L are most likely indicative of magnesium deficiency and those above 0.68mmol/L are hypermagnesemic. However, the meta-analysis also found that in patients with chronic diseases such as cardiovascular disease, type 2 diabetes, and hypertension associated with magnesium deficiency, the 95% confidence intervals for plasma ionized magnesium concentrations included those found for healthy individuals. This indicates, as noted for plasma total magnesium, that individuals in the designated adequate range for plasma ionized magnesium may also be magnesium deficient. Thus, additional measures of a dietary magnesium intake <250mg/day and a urinary magnesium excretion <80mg/day might help in identifying individuals who are very magnesium deficient even though they have plasma ionized magnesium concentrations in the low healthy reference range.

23c.8.3 Erythrocyte magnesium

Magnesium exists in erythrocytes in free, complex, and protein-bound forms. Magnesium concentrations in erythrocytes are about three times greater than those in plasma and reflect chronic rather than acute magnesium status because of the long lifespan of erythrocytes (120 days). Cellular magnesium is considered more reflective of magnesium status than circulating magnesium. However, erythrocyte magnesium concentrations have been reported as not correlating well with other tissue pools such as mononuclear blood cells or muscle (Ryzen et al., 1986; Elin, 1987). Thus, it is uncertain how well erythrocyte magnesium concentrations reflect body magnesium stores.

Determination of cellular magnesium concentrations takes special procedures including centrifugation to separate blood into an erythrocyte bottom layer, an intermediate buffy-coat layer of white blood cells, and a top layer of plasma. Erythrocytes are then acid digested and the magnesium content determined by atomic absorption spectrometry. Care must be taken to avoid hemolyzing the blood sample when obtaining the erythrocytes as it would result in loss of magnesium to the plasma portion. A reference range for the concentration of magnesium in erythrocytes of 1.65 to 2.65mmol/L has been reported (Costello & Nielsen, 2017). In depletion-repletion metabolic unit studies, a change in magnesium status determined by expressing erythrocyte magnesium as mmol/mg erythrocyte membrane protein appeared to be better than when expressed per erythrocyte cells, packed cells, or hemoglobin (Nielsen et al., 2003, 2007a, 2007b). The range in values found in these studies for erythrocyte membrane magnesium was 2.26 to 3.25 nmol/mg protein. However, this method of measuring status requires additional steps involving hemolyzing collected erythrocytes followed by washing and collecting the erythrocyte membranes after isolating them from the erythrocytes (Dodge et al., 1963).

The studies of Dodge and co-workers (Dodge et al., 1963) also determined erythrocyte ionized magnesium by hemolyzing the erythrocytes to release ionized magnesium for measurement by an ion selective electrode. However, determining the erythrocyte ionized magnesium content or expressing magnesium per cell or g hemoglobin did not appear to be any better for assessing cellular magnesium status than determining erythrocyte magnesium mmol/L packed cells. Further studies might find measuring

red blood cell membrane magnesium useful in the research setting.

Even though the assessment of body magnesium stores or CLMD may be slightly better using erythrocyte magnesium, the method of measurement is laborious and prone to error, that coupled with lack of a well validated reference range, preclude measurements of erythrocyte magnesium from being a useful clinical method for determining magnesium status. Erythrocyte membrane or packed cell magnesium determination may be useful in determining changes in magnesium status in long-term research studies of magnesium depletion and repletion.

23c.8.4 Magnesium in mononuclear and other cell types

Experimental findings from rats showing that changes in magnesium concentrations in mononuclear cells (lymphocytes and monocytes) reflected those in cardiac and skeletal muscle (Ryan & Ryan, 1979) was a stimulus to examine whether these cells could be used for magnesium status assessment in humans. Between 1980-2000, magnesium measurement of mononuclear cells was examined as indicators of magnesium status in a small number of studies involving human diseases associated with magnesium deficiency. Results from notable studies have been reviewed (Arnaud, 2008). Among the findings were the following: a significant correlation between magnesium concentrations in lymphocytes and muscle in patients with type 1 diabetes; lower lymphocyte ionized magnesium concentrations in migraine patients than in controls; lower lymphocyte total magnesium content in hypertensive patients than in controls; and a decrease in lymphocyte magnesium content in patients with congestive heart failure experiencing arrhythmias because of digitalis toxicity. Magnesium concentrations in platelets were also examined in patients with diabetes, obesity, and hypertension but were found not to be a more sensitive nor a more specific marker of magnesium status than other cellular magnesium measurements(Arnaud, 2008).

There is no reference range for cellular mononuclear magnesium. Only a few studies have indicated the normal concentration of magnesium in mononuclear cells. For example, a mean of 2.91±0.6fmol Mg/cell was obtained from 20 apparently healthy adults (Elin & Hosseini, 1985), whereas expressed as µmol/mg protein, mean cellular mononuclear magnesium concentrations ranged from 0.052 to 0.073µmol/mg protein (Elin & Johnson, 1982; Hosseini et al., 1983; Sjögren et al., 1986).

Although cellular mononuclear magnesium appears to be a better indicator of magnesium status or stores than serum total or ionized magnesium, several drawbacks preclude the use of cellular mononuclear magnesium as a routine clinical measure. For example, large blood samples are required for fractionating cellular components. Moreover, the fractionation method used to obtain the specific cell types is difficult, time consuming, and prone to errors. Analysis of magnesium concentrations in the specific cell types is usually performed by atomic absorption spectrometry.

23c.8.5 Urinary magnesium

Urinary magnesium excretion is an excellent indicator of magnesium intake and is usually measured by atomic absorption spectrometry. Urinary magnesium excretion data obtained from 27 metabolic unit studies have shown that when intakes were 200mg/day or lower, urinary magnesium excretion generally ranged from 40-80mg/day (1.65-3.29mmol/day). When dietary intakes were greater than 250mg/day, urinary magnesium excretion ranged from 80-160mg/day (3.29-6.58mmol/day) (Nielsen & Johnson, 2017).

Four depletion-repletion experiments involving metabolic unit studies with postmenopausal women found that urinary magnesium excretion was between 40-80mg/day (1.65-2.88mmol/day) with dietary magnesium intakes less than the EAR suggested for the United States and Canada of 175mg/day (Nielsen & Johnson, 2017). With magnesium intakes above 300mg/day, urinary magnesium excretion ranged from 100-140mg/day(4.11-5.76mmol/day) The four depletion-repletion experiments also found that urinary magnesium excretion responded quickly to changes in dietary magnesium. For example, in one study, women who consumed a mean magnesium intake of 390mg/day had a mean daily urinary excretion of120-130mg/day) (4.93-5.35mmol/day). When the women were switched to a diet supplying a mean magnesium intake of 120mg/day, mean urinary magnesium excretion decreased to 63mg (2.59mmol) per day within 6 days. A switch from a magnesium deficient intake to an adequate intake resulted in a similar change in urinary magnesium excretion within 6 days. These findings indicate that a single 24-hour urine sample or short-term determination of 24-hour urinary magnesium excretion alone is not useful as an indicator of current magnesium

status at the individual level because urinary magnesium excretion could be low while an individual still has an adequate magnesium status and vice versa. However, at the population level, urinary magnesium excretion could be an indicator of the amount of inadequate magnesium status in a population. This suggestion is supported by the findings that low urinary magnesium excretion has been associated with the risk of cardiovascular disease (Joosten et al., 2013a; Yamori et al., 2015) and hypertension (Joosten et al., 2013b).

Urinary magnesium excretion can also be useful in magnesium status assessment when combined with other measures of magnesium status. As indicated above, using a urinary magnesium excretion of < 80mg/day combined with a dietary magnesium intake < 250mg/day could increase the likelihood of identifying individuals who are magnesium-deficient even though serum magnesium concentrations are within the low healthy reference range. This suggestion is supported by a recent finding that young women with a mean urinary magnesium excretion of 58.2mg/day, a dietary magnesium intake of 145mg/day, and a normal serum magnesium responded to magnesium supplementation by increasing urinary magnesium excretion (Okamoto et al., 2023).

23c.8.6 Magnesium load test

The magnesium load test is considered by many to be the best method for determining total body magnesium status or general intracellular magnesium content. The basis for the magnesium load test is that the kidney regulates the body stores of magnesium by excreting excess magnesium when stores are adequate and retaining magnesium when stores are deficient. This test determines the percentage of magnesium retained over a given period time after parenteral administration of a magnesium load (Elin, 2010; Rude & Shils, 2006). Retention of a greater percentage of magnesium than that retained by individuals with adequate magnesium status indicates some depletion of body magnesium. The test depends upon normal renal function; it cannot be used on individuals with renal diseases or who are taking drugs that affect renal function. The magnesium load test has been used to identify magnesium deficiency in elderly individuals, and in conditions of chronic alcoholism, ischemic heart disease, hypertension, and Crohn's disease (Waters et al., 2008) where hypomagnesemia was not always found.

To perform this test properly, a basal urinary magnesium excretion is determined. Three consecutive 24-hour urine samples should be collected to yield the most accurate basal magnesium excretion and counteract the effects of both diurnal and random day-to-day variations in magnesium excretion. Nevertheless, measurement of a single 24-hour basal urinary magnesium excretion collected just before the magnesium load is most often used because it has been found to give acceptable results. After the determination of the basal urinary excretion, magnesium is administered parentally instead of orally to eliminate variability that would occur with intestinal absorption. A slow infusion, usually done over a period of 4 to 8 hours, is needed because plasma magnesium concentration determines the renal reabsorption threshold. A sudden increase in plasma magnesium above the normal level would increase urinary excretion and thus reduce magnesium retention, which could result in an erroneous assessment of status. Following the load, urinary excretion of magnesium is measured over the subsequent 24 to 48-hour period.

Although a 48-hour urine collection gives a slightly more accurate measure of magnesium retention, 24-hour urine collections have been found to correlate well with 48-hour urine collections (Gullestad et al., 1994) so are often used. Urine specimens should be collected with an acidifying agent to prevent precipitation of magnesium compounds at high pH. The net retention of magnesium is calculated by comparing basal magnesium excretion with net magnesium excretion after the load.

A standard protocol for administering the magnesium load has not been established. Examples of protocols reported include a 30mmol (720mg) of magnesium sulfate in 1000ml of isotonic saline for 8 hours at a rate of 125ml/hour (Gullestad et al., 1994); an intravenous infusion of 30 mmol (720mg) of magnesium in 500ml 5% dextrose over a period of 8-12 hours (Ismail et al., 2012) and an intravenous infusion of 0.2mmol/kg body weight (340mg for a 70kg individual) of magnesium sulfate in 250ml of 5% dextrose over 4 hours (Ozono et al., 1995). The percentage magnesium retention is calculated as:

[(Amount of Mg infused)
– (post-infusion urinary Mg)
– (baseline urinary Mg)] x 100%
divided by
(Amount of Mg infused)

(DiNicolantonio et al., 2018). The infusion protocols outlined above gave similar results in terms of suggesting the magnesium retention indicating magnesium adequacy and deficiency. A retention reference range providing the values indicating deficiency and adequacy has not been officially established. However, it has been suggested that a retention of < 20% indicates adequate

magnesium status; retention between > 20% to > 50% indicates borderline deficient magnesium status; and retention > 50% indicates magnesium deficiency (Costello & Nielsen, 2017). The borderline deficiency values, which would indicate CLMD, are apparently appropriate because individuals with values within this range reportedly had chronic diseases associated with a low magnesium status but did not exhibit hypomagnesemia (i.e., low serum magnesium concentrations).

The magnesium load test is invasive, time-consuming, and cumbersome; it requires close supervision for at least 24 hours after magnesium infusion to ensure oversight of both the total amount of magnesium infused and the urine collection. These drawbacks prevent the magnesium load test being a routine clinical measure of magnesium status. It has been used most often as a research tool to determine whether individuals with chronic disease are magnesium deficient.

23c.8.7 Muscle magnesium

Because muscle contains about 30% of the total body magnesium, muscle magnesium has been evaluated as a possible indicator of magnesium status. Very few studies have found a correlation between concentrations of magnesium in muscle and serum or erythrocytes (Elin, 1991). However, some significant correlations with mononuclear blood cells have been found in some disease states such as type 1 diabetes (Sjögren et al., 1986) and mild hypertension (Dyckner & Wester, 1985), which suggests that muscle magnesium may be a reliable indicator of intracellular magnesium status. This suggestion is supported by a metabolic unit study with postmenopausal women in which magnesium depletion by dietary means significantly decreased muscle magnesium, whereas repletion restored muscle magnesium levels (Lukaski & Nielsen, 2002).

In the magnesium depletion-repletion metabolic study outlined above, 40mg of muscle tissue was obtained by needle biopsy. After adipose and connective tissue were dissected from each specimen, the specimen was rinsed with distilled-deionized water and dried and digested with nitric acid. Magnesium was determined by inductively coupled plasma emission spectrometry. Muscle magnesium concentrations were 53.4, 48.1, and 51.6mmol/kg dry weight during the control, depletion, and repletion periods, respectively. However, the procedure to obtain muscle samples is extremely invasive and requires skill to obtain a sample with minimal fat and without causing a hematoma.

Proton nuclear magnetic resonance spectroscopy has also been used to measure ionized magnesium in skeletal muscle (Reyngoudt et al., 2019). However, this method of measurement is very expensive and requires the use of complex instruments and sophisticated software, which precludes its use in specialized laboratories only. Because other methods, such as the magnesium load test, and erythrocyte membrane and mononuclear cell magnesium determinations give status information of similar validity, using muscle magnesium to assess status in either the clinical or research setting is not useful.

23c.8.8 Buccal call magnesium

intracellular magnesium in sublingual epithelial cells was found to correlate with magnesium in atrial cell biopsies from open heart surgery patients(Haigney et al., 1995). Hence, the measurement of magnesium in sublingual epithelial cells has been suggested as a method that can assess CLMD(Silver, 2004). The cells are collected by gentle scratching of sublingual tissue and then fixed on a carbon slide with a cytology fixative. The magnesium in the fixed cells is determined by using energy-dispersive X-ray analysis. However, the test of magnesium in sublingual epithelial cells as a measure of magnesium status needs to be validated by using another test such as the magnesium-load test. Furthermore, the method is expensive and requires skill in collecting suitable sublingual epithelial cell samples for analysis, which has limited its use mainly to research laboratories.

23c.8.9 Magnesium status assessment methods of limited use

Magnesium balance (intake minus loss) is the principal measure used to determine dietary magnesium requirements. Most balance studies have been performed in clinical research centers. The method requires an accurate determination of magnesium from dietary intake and from the complete collection of urine and stool over at least 12 days. It also should include the collection and measurement of sweat, dermal, and menstrual losses of magnesium, although generally these losses have not been measured in reported balance studies. A negative balance (i.e., magnesium losses exceed magnesium intake) indicates a magnesium deficiency, and a positive balance (i.e., magnesium intake exceeds magnesium losses) indicates adequacy. Most reported balance studies prior to 1990 did not have well-controlled dietary intakes of magnesium, used short periods of data collection, and/or used less accurate methods to measure magnesium. A poorly controlled magnesium balance study was used to set the DRIs for magnesium in the United States and Canada (IOM, 1997). The most valid balance studies have involved individuals in a controlled metabolic unit environment who have been provided with a controlled and consistent diet (Hunt & Johnson, 2006; Nielsen, 2016). Such studies have indicated an EAR of 175mg/day for healthy 70kg individuals.

Fractional excretion of magnesium (FEMg) distinguishes between gastrointestinal and renal magnesium loss. This method has received some brief mentions as a measure of magnesium status (Ayuk & Gittoes, 2014; Costello & Nielsen, 2017). The following formula is used to determine this measure (Ayuk & Gittoes, 2014):

FEMg= [(uMg × sCr)/(sMg × uCr × 0.7)] × 100%

where uMg = urinary Mg concentration; sMg = serum Mg concentration; sCr= serum creatinine; uCr = urine creatinine

Fractional excretion of magnesium (FEMg) greater than 2% is considered an indication of renal wasting which implicates extra-renal magnesium losses, decreased intake, or decreased absorption of magnesium. As this method of assessment requires the determination of three variables, it provides no advantage to assess magnesium status over the use of serum magnesium, ionized magnesium, or the magnesium load test and hence has received limited application as a measure of magnesium status.

Magnesium depletion score (MDS) is an aggregate of several risk factors affecting the absorption and excretion of magnesium (Fan et al., 2021; Wang et al., 2022). The MDS is used to identify individuals with abnormal magnesium absorption and/or excretion that may result in a deficient magnesium status. The MDS is calculated by aggregating the factors:

1. Current use of diuretics counted as 1 point;
2. Current use of proton pump inhibitor counted as 1 point;
3. Heavy drinker (defined as > 1 drink/day for women and > 2 drinks/day for men) counted as 1 point;
4. Mildly decreased kidney function defined as estimated glomerular filtration rate (eGFR) 60mL/(min × 1.73m^2) < eGFR 90ml/min × 1.73m^2counted as 1 point;
5. Chronic kidney disease defined as eGFR < 60ml/min × 1.73m^2 counted as 2 points.

A score > 2 has been used to indicate magnesium deficiency associated with increased risk for systemic inflammation and cardiovascular mortality in adults (Fan et al., 2021). A score of > 3 combined with a dietary magnesium intake below the US RDA has been used to indicate magnesium deficiency associated with osteoporosis (Wang et al., 2022). This suggests that sample size or disease entity may influence the cut-point indicative of magnesium deficiency. This new suggested method of magnesium status assessment, especially for individuals with diseases associated with magnesium deficiency, needs further evaluation and validation before being accepted for general use.

This page titled 25.8: Indices of magnesium status (23c.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26: MAGNESIUM (CHAPTER 23C)

Abstract

Magnesium is a cofactor for over 600 enzymatic reactions vital for life and is a controlling factor in nerve transmission, skeletal and smooth muscle contraction, cardiac excitability, vasomotor tone, blood pressure, and bone turnover. Thus, magnesium deficiency can have detrimental consequences that include impaired physical and mental well-being and risk for chronic disease. Before 1990, a nutritional magnesium deficiency was considered rare and to be present mainly in clinical conditions associated with gastrointestinal or renal loss of magnesium. However, numerous reports of an association between a low magnesium intake and chronic diseases, especially those associated with chronic inflammatory stress, has led to the recognition that mild or subclinical magnesium deficiency, also known as chronic latent magnesium (CLMD) deficiency, may be quite prevalent. As a result, magnesium has become a nutrient of public health concern and a simple, rapid, and reliable clinical measure is needed to assess for magnesium deficiency. Numerous methods of magnesium status assessment have been developed, but all have shortcomings that impact their use in the clinical setting to accurately assess magnesium status.

The predominant method for assessing magnesium status at present is the determination of total serum or plasma magnesium. However, this method is not a sensitive indicator of body stores of magnesium. In the reference range of serum or plasma magnesium values, there is an interval in which individuals may have either CLMD or an adequate status. Combining measurement of total serum or plasma magnesium with both the determination of magnesium from dietary intake and urinary excretion appears to enhance the usefulness of this method to assess magnesium status.

Recent models of ion-selective electrodes have enabled the measurement of plasma ionized magnesium, the physiologically active form, to be acceptably accurate and precise using whole blood. As a result, this method is becoming more prevalent in the clinical setting because the procedure is simpler and requires less blood than the measurement of total magnesium in plasma or serum magnesium. However, plasma ionized magnesium has shortcomings similar to those for total plasma and serum magnesium regarding the sensitivity of the reference range to determine CLMD accurately.

Measurements of magnesium in erythrocytes, erythrocyte membranes, and mononuclear cells give a reasonable assessment of physiologically active body stores of magnesium. However, all these methods are laborious and prone to error, and hence only suitable for use in a research setting and not for routine clinical use.

Urinary magnesium excretion is an excellent indicator of magnesium intake. However, a single urinary magnesium determination may not reflect magnesium stores because it responds so rapidly to changes in dietary intake. Hence, it is mainly appropriate for use in population studies, or in combination with other measures of status.

The magnesium load test provides the most valid assessment of magnesium status. It has been used to identify magnesium deficiency in elderly individuals, in chronic alcoholism, and in several chronic diseases in which hypomagnesemia was not present. The test determines the percentage of urinary magnesium retained over a given period of time after parenteral administration of a magnesium load and yields an abbreviated balance determination. The test is invasive, time-consuming, and cumbersome; it requires close supervision for at least 24 hours after a magnesium load. Because of these drawbacks, it has been used mostly as a research tool.

Several other methods used to assess magnesium status also have drawbacks that preclude their use as routine clinical methods. Muscle magnesium determination is extremely invasive and requires skill to obtain suitable samples for analysis. Buccal cell magnesium determinations also require skill in obtaining cells, and special expensive equipment for analysis. Magnesium balance requires controlled and consistent dietary intakes and careful collection of urine and stool over a lengthy study period. Fractional excretion of magnesium requires the determination of three variables and has no apparent advantage over other well-established methods. The magnesium depletion score needs further evaluation and validation, although seems to be most useful for establishing status in individuals with a chronic disease. However, all these methods do not have an established validated reference range that indicates magnesium adequacy or deficiency.

At present, there is no simple, rapid, and reliable single clinical method to determine the presence of chronic, latent magnesium deficiency (CLMD). Determination of serum total or ionized magnesium remains the most acceptable choice in the clinical setting. The reliability of these measures of status could be improved if their values for the reference ranges indicative of deficient or adequate magnesium status are confirmed by other measures such as the determination of magnesium in both dietary intakes and urinary excretion.

26.1: Introduction (23c.1)
26.2: Functions of magnesium (23c.2)
26.3: Absorption and metabolism (23c.3)
26.4: Magnesium deficiency in humans (23c.4)
26.5: Food sources and dietary Intakes (23c.5)
26.6: Effects of high magnesium intakes (23c.6)
26.7: Occurrence of dietary deficiency (23c.7)
26.8: Indices of magnesium status (23c.8)

This page titled 26: Magnesium (Chapter 23c) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.1: Introduction (23c.1)

The human body contains about 760mg of magnesium at birth, 5g at age 4-5 months, and 25g when adult(World Health Organization and Food and Agriculture Organization of the United Nations, 2004). There are three body pools of magnesium (Rude & Shils, 2006). One pool has a turnover rate of less than 28hours. This pool of extracellular magnesium, which includes blood, contains 0.8-1.0% of body magnesium. The second pool has a turnover rate of about 11days. This pool is primarily intracellular magnesium, of which about 40% of the magnesium is found in soft tissues where it performs most of its essential functions. Skeletal magnesium is the third pool with a slow turnover rate measured in months or years in normal adults. Skeletal magnesium contains about 60% of body magnesium of which two-thirds is found within the hydration shell (bone mineral or poorly crystallized hydroxyapatite) and one-third on the crystal surface of the skeleton (Rude & Shils, 2006). Surface magnesium is readily exchangeable with serum and serves as a reservoir for the maintenance of magnesium for essential functions when intakes of magnesium are deficient.

This page titled 26.1: Introduction (23c.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.2: Functions of magnesium (23c.2)

Divalent magnesium ($Mg2^{+}$) is the fourth most abundant cation in the body and is second to potassium as the most abundant intracellular cation. Magnesium is a cofactor for over 600 enzymatic reactions vital to pathways including DNA, RNA, protein and adenosine 5'-triphosphate (ATP) synthesis; cellular energy production and storage; glycolysis; and cellular second messenger systems. The predominant enzymatic role of magnesium in these pathways involves an enzyme interacting with MgATP to complete a reaction. Magnesium also binds directly to enzymes to cause conformational change (allosteric activation) that gives a catalysis center for reaction to occur. Stabilization of ribonucleotides and deoxyribonucleotides by magnesium allow for DNA duplication, transcription, and maintenance, and transfer RNA function (Rude & Shils, 2006). Magnesium reacts with phosphates and carboxylates to stabilize membranes, affecting their fluidity and permeability which influences cellular ion channels, transporters, and signaling (Rude & Shils, 2006). Through regulating $Ca2^{+}$ and K^{+} movement in and out of the cell, magnesium is a controlling factor in nerve transmission, skeletal and smooth muscle contraction, cardiac excitability, vasomotor tone, blood pressure, and bone turnover.

This page titled 26.2: Functions of magnesium (23c.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.3: Absorption and metabolism (23c.3)

The jejunum and ileum are the primary sites of magnesium absorption; lesser amounts are absorbed by the cecum and colon. When dietary intakes of magnesium are near requirement, 45-55% is absorbed. With intakes less than 50% of requirement, the percentage of magnesium absorbed may increase to 65-70%. As dietary intake increases over requirement the percentage of magnesium absorbed decreases. For example, with an intake of 972mg, fractional absorption of magnesium was only 11% (Fine et al., 1991).

Magnesium is absorbed via both active transport (transcellular) and passive diffusion (paracellular). Active transport occurs in the ileum, although apparently also in the cecum and colon through the action of magnesium transporters: transient receptor potential melastatin (TRPM) type 6 (TRPM6) and TRPM7 (de Baaij et al., 2015). Passive paracellular diffusion accounts for about 90% magnesium absorption and is dependent on a high concentration in the jejunum and ileum and a transcellular potential difference generated by sodium transport (Romani, 2013).

Magnesium absorption can also be affected by other dietary components, including calcium, phosphorus, fiber, phytate, protein, and oxalates, all of which can decrease absorption, whereas vitamin D and its active metabolites, carbohydrates fermentable by bacteria, and oligosaccharides increase absorption (Nielsen, 2017).

The primary organ regulating magnesium homeostasis is the kidney where about 10% of total body magnesium is filtered daily through the glomeruli with the nephron recovering 95-99% of filtered magnesium. Passive diffusion associated with calcium, sodium, and water transport results in the reabsorption of about 10-25% of the magnesium (de Baaij et al., 2015). Passive diffusion in the thick ascending limb of the loop of Henle accounts for about 50-70% of the magnesium reabsorption. The remaining 5-10% of filtered magnesium is reabsorbed in the distal convoluted tubule via an active transcellular mechanism. The active transport channel in the kidney apparently involves TRPM6. When magnesium intakes are deficient, excretion of magnesium by the kidney decreases to 12-24mg/day to conserve body magnesium. Diets high in sodium, calcium and protein, and caffeine and alcohol consumption all enhance magnesium excretion (Costello & Rosanoff, 2020).

This page titled 26.3: Absorption and metabolism (23c.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.4: Magnesium deficiency in humans (23c.4)

Signs and symptoms of severe magnesium deficiency include positive Trousseau's and Chvostek's signs; muscle spasms, fasciculations, and tremor; and personality changes such as apathy, depression, nervousness, delirium, and hallucinations (Rude, 1998). However, severe magnesium deficiency involving these signs and symptoms is rare and usually associated with predisposing and complicating disease states that cause low magnesium intakes or impaired intestinal or renal absorption.

In conditions of long term moderate or mildly deficient intakes of magnesium or excessive excretion, magnesium is slowly released from bone to assure that vital essential functions are fulfilled. Nevertheless, there may still be a lack of magnesium required for some functions, but the signs associated with severe magnesium deficiency outlined above do not occur. This lack has been called chronic latent magnesium deficiency (CLMD)(Elin, 2010).

Chronic latent magnesium deficiency is considered a factor contributing to several chronic diseases, especially cardiovascular disease, including ischemic heart disease, sudden cardiac death, heart failure, atrial fibrillation, and ischemic stroke (Costello & Rosanoff, 2020). The associated risk of CLMD with cardiovascular disease is through impaired glucose metabolism and insulin action, dyslipidemia, hypertension, chronic inflammation, and impaired vasomotor tone (Costello & Rosanoff, 2020).

Chronic latent magnesium deficiency also may result in an increased risk for bone fractures and osteoporosis (Costello & Rosanoff, 2020). Apatite crystals formed in magnesium-deficient bone are increased in size, which makes bone more brittle, increasing the fracture risk. Magnesium deficiency can also reduce parathyroid hormone and 1,25-dihydroxycholecalciferol levels, which affect the regulation of magnesium-dependent calcium channels and changes in the release of inflammatory cytokines. Such effects enhance osteoclastic action and inhibit osteoblastic action, resulting in bone loss and an increased risk for osteoporosis.

Other chronic diseases with which CLMD has been associated include osteoarthritis, chondrocalcinosis, depression, and migraine headaches(Costello & Rosanoff, 2020). In addition, CLMD has been found to cause chronic inflammatory & oxidative stress, indicating that through these effects, CLMD may contribute to chronic diseases, other than those noted above, in which inflammatory or oxidative stress facilitates their occurrence.

This page titled 26.4: Magnesium deficiency in humans (23c.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.5: Food sources and dietary Intakes (23c.5)

Foods of plant origin provide about 50% of the magnesium intake for adults on mixed diets (Hunt & Meacham, 2001). Meats supply about 14-16% and dairy products (including milk) supply about 34% of the magnesium intakes in U.S. adults (Hunt & Meacham, 2001). Rich food sources of magnesium include whole grains / cereal, nuts, pulses, green leafy vegetables, and dark chocolate. Intermediate sources of magnesium include cheese, meats, and most seafoods. Refined grains are poor sources of magnesium.

This page titled 26.5: Food sources and dietary Intakes (23c.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.6: Effects of high magnesium intakes (23c.6)

Because the kidneys are very effective in the excretion of excessive magnesium, severe magnesium toxicity rarely occurs. When severe toxicity resulting in high blood magnesium occurs, the condition arises almost exclusively in individuals with kidney dysfunction or failure who are given magnesium salts or drugs such as laxatives or antacids. Symptoms of toxicity include lethargy, confusion, nausea, diarrhea, impaired breathing, hypotension, muscle weakness, and heart arrhythmias (Romani, 2013). Signs of hypermagnesemia also include low blood calcium and high blood potassium. The heart rhythm changes in magnesium toxicity are caused by abnormal electrical conduction which occurs at the nervous, muscular, and cardiac level.

Severe magnesium toxicity caused by excessive intake through food has not been reported. High intakes of supplemental magnesium available in forms such as aspartate, citrate, chloride, gluconate, lactate, and oxide can cause adverse effects such as diarrhea, abdominal cramping, and nausea. A tolerable upper intake level (UL), defined as the highest level of daily intake that is likely to pose no risks of adverse health effects to almost all individuals in the general population, was set at 350mg of supplemental magnesium in the United States and Canada for children older than 8 years and adults by the (Institute of Medicine. 1997),

This page titled 26.6: Effects of high magnesium intakes (23c.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.7: Occurrence of dietary deficiency (23c.7)

In the official Dietary Reference Intakes (DRIs) established for the United States and Canada (Institute of Medicine. 1997), the Estimated Average Requirements, (EAR, defined as average daily intake estimated to meet the daily requirements of 50% of healthy individuals) were set for individuals from aged 19 to 70 years and over at 330 to 350mg/day for males and at 255 to 265mg/day for females. These EAR values for adults are consistent with those set for adults in several other countries, as shown in Table 23c.1.

Table 23c.1 Comparison of dietary requirements of magnesium for adults (mg/day) set by four different authorities. A panel from the European Food Safety Authority (2015). considered that Average Requirements and Population Reference Intakes for magnesium cannot be derived for adults, infants or children, and therefore defined Adequate Intakes (AIs), based on observed intakes in healthy populations in the European Union (EU). EAR, Estimated Average Requirement; RDA, Recommended Dietary Allowance; RNI, Reference Nutrient Intake; RDI, Recommended Dietary Intake; RNI, Recommended Nutrient Intake; NA, Not Available.*

US/Canada Institute of Medicine (1997).	Age(y)	EAR	RDA
Males	19-30	330	400
	31-70+	350	420
Females	19-30	255	310
	31-70+	265	320
United Kingdom Dietary Reference Values (1991).	Age	EAR	RNI
Males	19-75+	250	300
Females	19-75+	200	270
Australia/New Zealand NZ Ministry of Health (2006).	Age	EAR	RDI
Males	19-30	330	400
	31-75+	350	420
Females	19-30	255	310
	31-75+	265	320
WHO (2004).	Age		RNI
Males	19-65	NA	260
	65+	NA	224
Females	19-65	NA	220
	65+	NA	190

In a United States survey performed in 2005-2006, magnesium intakes less than the EAR were reported in 48% of females in ages 31-50y, 55% in ages 51-70y, and 70% in ages 71y, and over; for males, intakes less than the EAR were 45% in ages 31-50y, 58% in ages 51-70y and 80% in ages 71y and over (Moshfegh et al., 2009).

Recently, it has been suggested that the DRIs for magnesium established for the United States and Canada should be revised and lowered to reflect findings of the improved balance study data (Nielsen, 2016). These balance study data indicated that the EAR and the Recommended Dietary Allowance (RDA, defined as average daily intake sufficient to meet the daily requirements

for almost all (about 97%) of healthy individuals) for 70kg healthy adults should be 175 and 250mg/d for 70kg healthy male and female individuals, and increase or decrease based on body weight, independent of age. The balance data found neutral balance (i.e., when intake and loss of magnesium from the body were equal) at 2.36mg/kg/day, the value indicating the amount of decrease or increase in the DRIs with each change in kg of body weight. Even with the lower DRIs, many individuals weigh more than 70kg and/or have health problems which will result in an increase in the need for magnesium. As a consequence, it is likely that over 25% of the United States population have an intake of magnesium less than the EAR. Thus, because mild or moderate magnesium deficiency can increase the risk for numerous chronic disorders, it is apparent that magnesium is a nutrient of concern requiring a clinically useful and sound indicator of magnesium status.

This page titled 26.7: Occurrence of dietary deficiency (23c.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

26.8: Indices of magnesium status (23c.8)

Because magnesium has so many critical functions, the body has strong mechanisms, such as reducing urinary excretion and releasing skeletal stores of magnesium, to assure that magnesium is readily available to fulfill those functions. In addition, mild or moderate magnesium deficiency is mostly asymptomatic. As a result of these characteristics, success in finding a cost-effective, time-efficient, and reliable indicator of magnesium status is lacking. Among the status indicators that have been evaluated are serum or plasma total magnesium, serum ionized magnesium, erythrocyte magnesium, erythrocyte membrane magnesium, blood mononuclear cell magnesium, urinary magnesium excretion, magnesium load test, sublingual magnesium, muscle magnesium, and fractional excretion of magnesium. An additional factor that may be useful in evaluating exposure to magnesium is dietary intake.

23c.8.1 Serum/plasma magnesium

At present, total serum magnesium is the predominant clinical test to assess magnesium status, although plasma total magnesium is sometimes measured instead of serum magnesium. However, serum is preferred because anti-coagulants might be contaminated with magnesium or affect assessment methods. Serum magnesium determination is inexpensive and easily performed using atomic absorption spectrophotometry or colorimetric methods. The normal range for serum magnesium is 0.75-0.95mmol/L. (Figure 23c.1). Hence, theoretically, serum concentrations above 0.75mmol/L should be considered magnesium sufficient, and those below 0.75mmol/L should be considered magnesium deficient, although, as indicated below, perhaps not rightly so (Costello et al., 2016). Serum magnesium concentrations are reportedly acceptable for determining a severe lack of magnesium intake or an elevated urinary excretion of magnesium resulting in a depletion of body stores. This finding is based on the results of a study in which exposure to extreme dietary magnesium deprivation (less than 12mg/day) resulted in a decline in serum magnesium from 0.80mmol/L to 0.61mmol/L in three weeks (Fatemi et al., 1991). However, in mild or moderate dietary magnesium deprivation, serum magnesium concentrations can remain in the normal range after several weeks, based on controlled metabolic unit studies of postmenopausal women whose dietary intakes were changed from adequacy (250mg/day) to moderate deficiency (80-160mg/day) (Nielsen & Johnson, 2017). Nevertheless, the magnesium-deprived women showed signs indicative of magnesium deficiency that responded to magnesium supplementation, indicating their body tissues were becoming magnesium deficient. Results of these metabolic unit studies are consistent with clinical studies of patients with diabetes mellitus, alcoholism, and malabsorption syndromes, all of whom had serum magnesium concentrations in the sufficient range, yet had low magnesium concentrations in erythrocytes, mononuclear cells, and muscle(Ryzen et al., 1986; Nadler et al., 1992; Abbott et al., 1994; Rude & Olerich, 1996). These findings indicate that individuals with serum/plasma magnesium concentrations within the current reference range might be deficient enough to have a bearing on the risk for chronic disease.

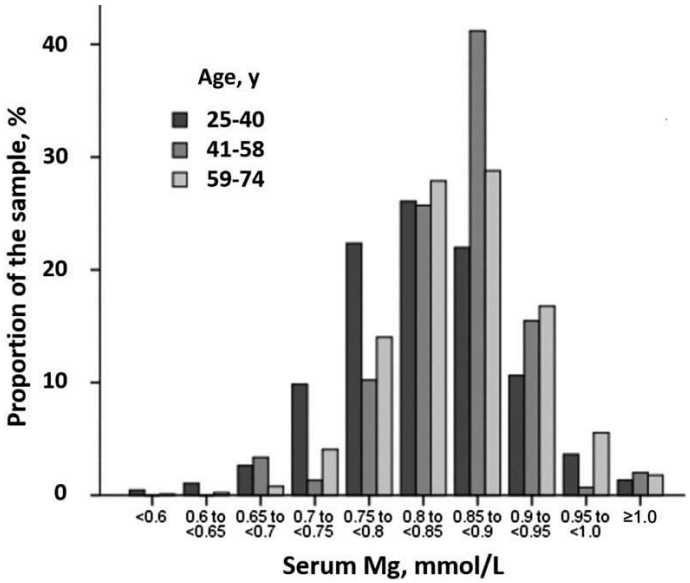


Figure 23c.1 Age-specific distributions of serum magnesium in US adults. Data were derived from NHANES I (1971 1975). Modified from (Costello et al. 2016).

Recently, it has been suggested that the reference range for serum magnesium should indicate that some individuals with serum magnesium concentrations between 0.75mmol/L and 0.85mmol/L might have CLMD(Costello et al., 2016). A review (Elin, 2010) has indicated that magnesium deficiency is present in 90% of individuals with serum magnesium concentrations of 0.70mmol/L or less, in 50% of individuals with concentrations of 0.75mmol/L, and in 10% of individuals with concentrations of 0.85mmol/L, emphasizing that serum magnesium alone lacks reliability as a status indicator of CLMD. Moreover, it also indicates that individuals with low serum magnesium concentrations might have adequate magnesium status.

The reliability of serum magnesium to indicate mild or moderate magnesium deficiency may be improved by also assessing urinary magnesium excretion and dietary magnesium intake. Metabolic unit findings suggest that individuals with serum magnesium concentrations >0.75mmol/L or as high as 0.85mmol/L could have a magnesium deficit such that they respond to magnesium supplementation if they have a dietary history showing a magnesium intake <250mg/day and a urinary excretion of magnesium of <80mg/day (Nielsen & Johnson, 2017).

Factors affecting serum magnesium

There also is a need to appreciate the numerous factors affecting serum magnesium concentrations when considering the reliability of this as a measure of magnesium status.

Diurnal variation in serum magnesium concentrations occurs: values are lower in the morning than in the evening (Wilimzig et al., 1996).

Strenuous exercise reduces serum magnesium through a shift from plasma into erythrocytes and an increase in excretion through urinary excretion and sweat loss (Lukaski, 2000).

Various drugs, such as diuretics, proton pump inhibitors, cancer therapies, and antibiotics may decrease serum magnesium concentrations (Rude, 1998), whereas antacids and cathartics may elevate concentrations (Arnaud MJ, 2008).

Disease states such as osteoporosis may decrease concentrations (Stendig-Lindberg et al., 1993). Chronic renal failure may increase concentrations by impairing magnesium homeostasis by the kidney (Elin, 1991).

Serum albumin concentrations are related to serum magnesium in a linear fashion (Kroll & Elin, 1985). Thus, hypoalbuminemia may result in low serum magnesium concentrations, and hyperalbuminemia in high serum magnesium concentrations.

23c.8.2 Serum, plasma, and blood ionized magnesium

Ionized magnesium is the physiologically active form of magnesium that is transported across cell membranes and participates in metabolic interactions and enzymatic reactions. Thus, it has received attention as a better measure than serum total magnesium for indicating the status of magnesium for essential cellular functions in the body. Ionized magnesium is determined in serum, plasma, or whole blood samples by using equipment with ion-selective electrodes. The first ion-selective electrodes were challenging to use, cost-intensive, and had poor ability to discriminate ionized magnesium from ionized calcium; this inhibited the use of ionized magnesium as a status-assessment measure. However, recent models of ion-selective electrodes are cost-effective and have been shown to be acceptably accurate and precise in measuring only ionized magnesium.

Originally, serum ionized magnesium was often determined, but at present, it is plasma magnesium that is mostly measured in whole blood; the ionized magnesium in erythrocytes is not measured. An advantage of using whole blood is that less blood and fewer procedures are needed for analyses. High pH and high concentrations of ionized calcium, lipophilic cations, and thiocyanate (caused by smoking tobacco) can interfere with the determination of plasma ionized magnesium by ion-selective electrodes (Ansu Baidoo et al., 2023).

Plasma ionized magnesium as a status assessment measure has the same problems as those outlined for plasma total magnesium; the body also attempts to maintain magnesium concentrations in plasma at a level needed for vital functions through reduced urinary excretion and bone resorption. Factors that affect plasma total magnesium, such as diuretics, proton pump inhibitors, and various disease states, also affect plasma ionized magnesium. An additional issue is that a reference range for plasma ionized magnesium indicative of adequate magnesium status has not been firmly established and appears to vary depending upon the instrument used for determining a range. A systematic review and meta-analysis involving all types of instruments used found a 95% confidence interval for ionized magnesium in healthy adults to be 0.40 to 0.68mmol/L (Ansu Baidoo et al., 2023). Using this as a reference range indicates plasma ionized magnesium concentrations below 0.40mmol/L are most likely indicative of magnesium deficiency and those above 0.68mmol/L are hypermagnesemic. However, the meta-analysis also found that in patients with chronic diseases such as cardiovascular disease, type 2 diabetes, and hypertension associated with magnesium deficiency, the 95% confidence intervals for plasma ionized magnesium concentrations included those found for healthy individuals. This indicates, as noted for plasma total magnesium, that individuals in the designated adequate range for plasma ionized magnesium may also be magnesium deficient. Thus, additional measures of a dietary magnesium intake <250mg/day and a urinary magnesium excretion <80mg/day might help in identifying individuals who are very magnesium deficient even though they have plasma ionized magnesium concentrations in the low healthy reference range.

23c.8.3 Erythrocyte magnesium

Magnesium exists in erythrocytes in free, complex, and protein-bound forms. Magnesium concentrations in erythrocytes are about three times greater than those in plasma and reflect chronic rather than acute magnesium status because of the long lifespan of erythrocytes (120 days). Cellular magnesium is considered more reflective of magnesium status than circulating magnesium. However, erythrocyte magnesium concentrations have been reported as not correlating well with other tissue pools such as mononuclear blood cells or muscle (Ryzen et al., 1986; Elin, 1987). Thus, it is uncertain how well erythrocyte magnesium concentrations reflect body magnesium stores.

Determination of cellular magnesium concentrations takes special procedures including centrifugation to separate blood into an erythrocyte bottom layer, an intermediate buffy-coat layer of white blood cells, and a top layer of plasma. Erythrocytes are then acid digested and the magnesium content determined by atomic absorption spectrometry. Care must be taken to avoid hemolyzing the blood sample when obtaining the erythrocytes as it would result in loss of magnesium to the plasma portion. A reference range for the concentration of magnesium in erythrocytes of 1.65 to 2.65mmol/L has been reported (Costello & Nielsen, 2017). In depletion-repletion metabolic unit studies, a change in magnesium status determined by expressing erythrocyte magnesium as mmol/mg erythrocyte membrane protein appeared to be better than when expressed per erythrocyte cells, packed cells, or hemoglobin (Nielsen et al., 2003, 2007a, 2007b). The range in values found in these studies for erythrocyte membrane magnesium was 2.26 to 3.25 nmol/mg protein. However, this method of measuring status requires additional steps involving hemolyzing collected erythrocytes followed by washing and collecting the erythrocyte membranes after isolating them from the erythrocytes (Dodge et al., 1963).

The studies of Dodge and co-workers (Dodge et al., 1963) also determined erythrocyte ionized magnesium by hemolyzing the erythrocytes to release ionized magnesium for measurement by an ion selective electrode. However, determining the erythrocyte ionized magnesium content or expressing magnesium per cell or g hemoglobin did not appear to be any better for assessing cellular magnesium status than determining erythrocyte magnesium mmol/L packed cells. Further studies might find measuring

red blood cell membrane magnesium useful in the research setting.

Even though the assessment of body magnesium stores or CLMD may be slightly better using erythrocyte magnesium, the method of measurement is laborious and prone to error, that coupled with lack of a well validated reference range, preclude measurements of erythrocyte magnesium from being a useful clinical method for determining magnesium status. Erythrocyte membrane or packed cell magnesium determination may be useful in determining changes in magnesium status in long-term research studies of magnesium depletion and repletion.

23c.8.4 Magnesium in mononuclear and other cell types

Experimental findings from rats showing that changes in magnesium concentrations in mononuclear cells (lymphocytes and monocytes) reflected those in cardiac and skeletal muscle (Ryan & Ryan, 1979) was a stimulus to examine whether these cells could be used for magnesium status assessment in humans. Between 1980-2000, magnesium measurement of mononuclear cells was examined as indicators of magnesium status in a small number of studies involving human diseases associated with magnesium deficiency. Results from notable studies have been reviewed (Arnaud, 2008). Among the findings were the following: a significant correlation between magnesium concentrations in lymphocytes and muscle in patients with type 1 diabetes; lower lymphocyte ionized magnesium concentrations in migraine patients than in controls; lower lymphocyte total magnesium content in hypertensive patients than in controls; and a decrease in lymphocyte magnesium content in patients with congestive heart failure experiencing arrhythmias because of digitalis toxicity. Magnesium concentrations in platelets were also examined in patients with diabetes, obesity, and hypertension but were found not to be a more sensitive nor a more specific marker of magnesium status than other cellular magnesium measurements (Arnaud, 2008).

There is no reference range for cellular mononuclear magnesium. Only a few studies have indicated the normal concentration of magnesium in mononuclear cells. For example, a mean of 2.91±0.6fmol Mg/cell was obtained from 20 apparently healthy adults (Elin & Hosseini, 1985), whereas expressed as µmol/mg protein, mean cellular mononuclear magnesium concentrations ranged from 0.052 to 0.073µmol/mg protein (Elin & Johnson, 1982; Hosseini et al., 1983; Sjögren et al., 1986).

Although cellular mononuclear magnesium appears to be a better indicator of magnesium status or stores than serum total or ionized magnesium, several drawbacks preclude the use of cellular mononuclear magnesium as a routine clinical measure. For example, large blood samples are required for fractionating cellular components. Moreover, the fractionation method used to obtain the specific cell types is difficult, time consuming, and prone to errors. Analysis of magnesium concentrations in the specific cell types is usually performed by atomic absorption spectrometry.

23c.8.5 Urinary magnesium

Urinary magnesium excretion is an excellent indicator of magnesium intake and is usually measured by atomic absorption spectrometry. Urinary magnesium excretion data obtained from 27 metabolic unit studies have shown that when intakes were 200mg/day or lower, urinary magnesium excretion generally ranged from 40-80mg/day (1.65-3.29mmol/day). When dietary intakes were greater than 250mg/day, urinary magnesium excretion ranged from 80-160mg/day (3.29-6.58mmol/day) (Nielsen & Johnson, 2017).

Four depletion-repletion experiments involving metabolic unit studies with postmenopausal women found that urinary magnesium excretion was between 40-80mg/day (1.65-2.88mmol/day) with dietary magnesium intakes less than the EAR suggested for the United States and Canada of 175mg/day (Nielsen & Johnson, 2017). With magnesium intakes above 300mg/day, urinary magnesium excretion ranged from 100-140mg/day(4.11-5.76mmol/day) The four depletion-repletion experiments also found that urinary magnesium excretion responded quickly to changes in dietary magnesium. For example, in one study, women who consumed a mean magnesium intake of 390mg/day had a mean daily urinary excretion of120-130mg/day) (4.93-5.35mmol/day). When the women were switched to a diet supplying a mean magnesium intake of 120mg/day, mean urinary magnesium excretion decreased to 63mg (2.59mmol) per day within 6 days. A switch from a magnesium deficient intake to an adequate intake resulted in a similar change in urinary magnesium excretion within 6 days. These findings indicate that a single 24-hour urine sample or short-term determination of 24-hour urinary magnesium excretion alone is not useful as an indicator of current magnesium

status at the individual level because urinary magnesium excretion could be low while an individual still has an adequate magnesium status and vice versa. However, at the population level, urinary magnesium excretion could be an indicator of the amount of inadequate magnesium status in a population. This suggestion is supported by the findings that low urinary magnesium excretion has been associated with the risk of cardiovascular disease (Joosten et al., 2013a; Yamori et al., 2015) and hypertension (Joosten et al., 2013b).

Urinary magnesium excretion can also be useful in magnesium status assessment when combined with other measures of magnesium status. As indicated above, using a urinary magnesium excretion of < 80mg/day combined with a dietary magnesium intake < 250mg/day could increase the likelihood of identifying individuals who are magnesium-deficient even though serum magnesium concentrations are within the low healthy reference range. This suggestion is supported by a recent finding that young women with a mean urinary magnesium excretion of 58.2mg/day, a dietary magnesium intake of 145mg/day, and a normal serum magnesium responded to magnesium supplementation by increasing urinary magnesium excretion(Okamoto et al., 2023).

23c.8.6 Magnesium load test

The magnesium load test is considered by many to be the best method for determining total body magnesium status or general intracellular magnesium content. The basis for the magnesium load test is that the kidney regulates the body stores of magnesium by excreting excess magnesium when stores are adequate and retaining magnesium when stores are deficient. This test determines the percentage of magnesium retained over a given period time after parenteral administration of a magnesium load (Elin, 2010; Rude & Shils, 2006). Retention of a greater percentage of magnesium than that retained by individuals with adequate magnesium status indicates some depletion of body magnesium. The test depends upon normal renal function; it cannot be used on individuals with renal diseases or who are taking drugs that affect renal function. The magnesium load test has been used to identify magnesium deficiency in elderly individuals, and in conditions of chronic alcoholism, ischemic heart disease, hypertension, and Crohn's disease (Waters et al., 2008) where hypomagnesemia was not always found.

To perform this test properly, a basal urinary magnesium excretion is determined. Three consecutive 24-hour urine samples should be collected to yield the most accurate basal magnesium excretion and counteract the effects of both diurnal and random day-to-day variations in magnesium excretion. Nevertheless, measurement of a single 24-hour basal urinary magnesium excretion collected just before the magnesium load is most often used because it has been found to give acceptable results. After the determination of the basal urinary excretion, magnesium is administered parentally instead of orally to eliminate variability that would occur with intestinal absorption. A slow infusion, usually done over a period of 4 to 8 hours, is needed because plasma magnesium concentration determines the renal reabsorption threshold. A sudden increase in plasma magnesium above the normal level would increase urinary excretion and thus reduce magnesium retention, which could result in an erroneous assessment of status. Following the load, urinary excretion of magnesium is measured over the subsequent 24 to 48-hour period.

Although a 48-hour urine collection gives a slightly more accurate measure of magnesium retention, 24-hour urine collections have been found to correlate well with 48-hour urine collections (Gullestad et al., 1994) so are often used. Urine specimens should be collected with an acidifying agent to prevent precipitation of magnesium compounds at high pH. The net retention of magnesium is calculated by comparing basal magnesium excretion with net magnesium excretion after the load.

A standard protocol for administering the magnesium load has not been established. Examples of protocols reported include a 30mmol (720mg) of magnesium sulfate in 1000ml of isotonic saline for 8 hours at a rate of 125ml/hour (Gullestad et al., 1994); an intravenous infusion of 30 mmol (720mg) of magnesium in 500ml 5% dextrose over a period of 8-12 hours (Ismail et al., 2012) and an intravenous infusion of 0.2mmol/kg body weight (340mg for a 70kg individual) of magnesium sulfate in 250ml of 5% dextrose over 4 hours (Ozono et al., 1995). The percentage magnesium retention is calculated as:

[(Amount of Mg infused)
– (post-infusion urinary Mg)
– (baseline urinary Mg)] x 100%
divided by
(Amount of Mg infused)

(DiNicolantonio et al., 2018). The infusion protocols outlined above gave similar results in terms of suggesting the magnesium retention indicating magnesium adequacy and deficiency. A retention reference range providing the values indicating deficiency and adequacy has not been officially established. However, it has been suggested that a retention of < 20% indicates adequate

magnesium status; retention between > 20% to > 50% indicates borderline deficient magnesium status; and retention > 50% indicates magnesium deficiency (Costello & Nielsen, 2017). The borderline deficiency values, which would indicate CLMD, are apparently appropriate because individuals with values within this range reportedly had chronic diseases associated with a low magnesium status but did not exhibit hypomagnesemia (i.e., low serum magnesium concentrations).

The magnesium load test is invasive, time-consuming, and cumbersome; it requires close supervision for at least 24 hours after magnesium infusion to ensure oversight of both the total amount of magnesium infused and the urine collection. These drawbacks prevent the magnesium load test being a routine clinical measure of magnesium status. It has been used most often as a research tool to determine whether individuals with chronic disease are magnesium deficient.

23c.8.7 Muscle magnesium

Because muscle contains about 30% of the total body magnesium, muscle magnesium has been evaluated as a possible indicator of magnesium status. Very few studies have found a correlation between concentrations of magnesium in muscle and serum or erythrocytes (Elin, 1991). However, some significant correlations with mononuclear blood cells have been found in some disease states such as type 1 diabetes (Sjögren et al., 1986) and mild hypertension (Dyckner & Wester, 1985), which suggests that muscle magnesium may be a reliable indicator of intracellular magnesium status. This suggestion is supported by a metabolic unit study with postmenopausal women in which magnesium depletion by dietary means significantly decreased muscle magnesium, whereas repletion restored muscle magnesium levels (Lukaski & Nielsen, 2002).

In the magnesium depletion-repletion metabolic study outlined above, 40mg of muscle tissue was obtained by needle biopsy. After adipose and connective tissue were dissected from each specimen, the specimen was rinsed with distilled-deionized water and dried and digested with nitric acid. Magnesium was determined by inductively coupled plasma emission spectrometry. Muscle magnesium concentrations were 53.4, 48.1, and 51.6mmol/kg dry weight during the control, depletion, and repletion periods, respectively. However, the procedure to obtain muscle samples is extremely invasive and requires skill to obtain a sample with minimal fat and without causing a hematoma.

Proton nuclear magnetic resonance spectroscopy has also been used to measure ionized magnesium in skeletal muscle (Reyngoudt et al., 2019). However, this method of measurement is very expensive and requires the use of complex instruments and sophisticated software, which precludes its use in specialized laboratories only. Because other methods, such as the magnesium load test, and erythrocyte membrane and mononuclear cell magnesium determinations give status information of similar validity, using muscle magnesium to assess status in either the clinical or research setting is not useful.

23c.8.8 Buccal call magnesium

intracellular magnesium in sublingual epithelial cells was found to correlate with magnesium in atrial cell biopsies from open heart surgery patients(Haigney et al., 1995). Hence, the measurement of magnesium in sublingual epithelial cells has been suggested as a method that can assess CLMD(Silver, 2004). The cells are collected by gentle scratching of sublingual tissue and then fixed on a carbon slide with a cytology fixative. The magnesium in the fixed cells is determined by using energy-dispersive X-ray analysis. However, the test of magnesium in sublingual epithelial cells as a measure of magnesium status needs to be validated by using another test such as the magnesium-load test. Furthermore, the method is expensive and requires skill in collecting suitable sublingual epithelial cell samples for analysis, which has limited its use mainly to research laboratories.

23c.8.9 Magnesium status assessment methods of limited use

Magnesium balance (intake minus loss) is the principal measure used to determine dietary magnesium requirements. Most balance studies have been performed in clinical research centers. The method requires an accurate determination of magnesium from dietary intake and from the complete collection of urine and stool over at least 12 days. It also should include the collection and measurement of sweat, dermal, and menstrual losses of magnesium, although generally these losses have not been measured in reported balance studies. A negative balance (i.e., magnesium losses exceed magnesium intake) indicates a magnesium deficiency, and a positive balance (i.e., magnesium intake exceeds magnesium losses) indicates adequacy. Most reported balance studies prior to 1990 did not have well-controlled dietary intakes of magnesium, used short periods of data collection, and/or used less accurate methods to measure magnesium. A poorly controlled magnesium balance study was used to set the DRIs for magnesium in the United States and Canada (IOM, 1997). The most valid balance studies have involved individuals in a controlled metabolic unit environment who have been provided with a controlled and consistent diet (Hunt & Johnson, 2006; Nielsen, 2016). Such studies have indicated an EAR of 175mg/day for healthy 70kg individuals.

Fractional excretion of magnesium (FEMg) distinguishes between gastrointestinal and renal magnesium loss. This method has received some brief mentions as a measure of magnesium status (Ayuk & Gittoes, 2014; Costello & Nielsen, 2017). The following formula is used to determine this measure (Ayuk & Gittoes, 2014):

FEMg= [(uMg × sCr)/(sMg × uCr × 0.7)] × 100%

where uMg = urinary Mg concentration; sMg = serum Mg concentration; sCr= serum creatinine; uCr = urine creatinine

Fractional excretion of magnesium (FEMg) greater than 2% is considered an indication of renal wasting which implicates extra-renal magnesium losses, decreased intake, or decreased absorption of magnesium. As this method of assessment requires the determination of three variables, it provides no advantage to assess magnesium status over the use of serum magnesium, ionized magnesium, or the magnesium load test and hence has received limited application as a measure of magnesium status.

Magnesium depletion score (MDS) is an aggregate of several risk factors affecting the absorption and excretion of magnesium (Fan et al., 2021; Wang et al., 2022). The MDS is used to identify individuals with abnormal magnesium absorption and/or excretion that may result in a deficient magnesium status. The MDS is calculated by aggregating the factors:

1. Current use of diuretics counted as 1 point;
2. Current use of proton pump inhibitor counted as 1 point;
3. Heavy drinker (defined as > 1 drink/day for women and > 2 drinks/day for men) counted as 1 point;
4. Mildly decreased kidney function defined as estimated glomerular filtration rate (eGFR) 60mL/(min × 1.73m^2) < eGFR 90ml/min × 1.73m^2counted as 1 point;
5. Chronic kidney disease defined as eGFR < 60ml/min × 1.73m^2 counted as 2 points.

A score > 2 has been used to indicate magnesium deficiency associated with increased risk for systemic inflammation and cardiovascular mortality in adults (Fan et al., 2021). A score of > 3 combined with a dietary magnesium intake below the US RDA has been used to indicate magnesium deficiency associated with osteoporosis (Wang et al., 2022). This suggests that sample size or disease entity may influence the cut-point indicative of magnesium deficiency. This new suggested method of magnesium status assessment, especially for individuals with diseases associated with magnesium deficiency, needs further evaluation and validation before being accepted for general use.

This page titled 26.8: Indices of magnesium status (23c.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27: COPPER (CHAPTER 24B)

Abstract

The role of copper in human health and well-being is considered, starting with copper metabolism, from uptake across the gut, to the liver and the incorporation into cupro-enzymes. Food sources of copper are noted, and the amount required to maintain health at different stages of the life course given. The consequences of copper deficiency are described, with data largely obtained from inherited diseases of copper metabolism and from the rare cases of copper overload; the effects of high copper are also discussed. Markers of copper status are assessed. Low serum copper levels are a marker of severe copper deficiency, but do not identify marginal copper deficiency Many factors, including infection and stress, modify serum copper levels, emphasising the unreliability of the indicator. The majority of serum copper is bound to ceruloplasmin, and parameters that affect ceruloplasmin are also discussed. Other possible markers of copper status include cuproenzymes such as Cu, Zn superoxide dismutase, cytochrome oxidase, lysyl oxidase and diamine oxidase, but they are difficult to measure, and normal ranges have not been established. When inside cells, copper is bound to small proteins called chaperones. mRNA or protein levels of these chaperones have been proposed as markers of status, with varying degrees of acceptability. Other markers of copper status, such as hair or nail copper levels, are also considered, but are not ideal and can be altered by factors other than copper status.

27.1: Copper (24b.0)

27.2: Serum copper (24b.1)

27.3: Serum ceruloplasmin (24b.2)

27.4: Erythrocyte superoxide dismutase (24b.3)

27.5: Cytochrome c oxidase (24b.4)

27.6: Skin lysyl oxidase (24b.5)

27.7: Diamine oxidase (24b.6)

27.8: Copper chaperones (24b.7)

27.9: Other indices of copper status (24b.8)

This page titled 27: Copper (Chapter 24b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.1: Copper (24b.0)

The adult human body contains 70–80mg of copper. Of this, approximately 25% is in the skeletal muscle, 15% in the skin, 15% in the bone marrow, 19% in the skeleton, 8.0 to 15% in the liver, and 8.0% in the brain. The liver and brain contain the highest concentrations of copper. Like iron, copper can be found in two valences, Cu^{+} and Cu^{2+}, and it is this ability to gain and lose electrons that give the enzymes into which it is incorporated their function, catalyzing oxidation-reduction reactions. Examples of copper enzymes (cupro-enzymes) are given in (Table 24b.1).

Table 24b.1 Function of some copper enzymes with oxidation and reduction activity in humans. From Uauy et al., American Journal c 1998

Copper Enzyme	Function
Ceruloplasmin (Ferroxidase I)	Copper transport, iron mobilization, antioxidan
Cytochrome c oxidase	Electron transport, terminal oxidase
Superoxide dismutase	Free radical detoxication
Tyrosinase	Synthesis of melanin
Lysyl oxidase	Collagen and elastin cross-linking
Amine oxidases	Deamination of primary amines
Dopamine-β-monooxygenase	Dopamine → norepi-nephrine
Peptidylglycine monooxygenase	α-Amidation of neuro peptides

• Functions of copper

Many clinical features of copper deficiency can be explained by changes in the activities of cuproenzymes. For instance, the ferroxidases, of which there are several examples, are involved in oxidising Fe^{2+} to Fe^{3+} to facilitate iron incorporation into trans-ferrin. Ceruloplasmin, a serum protein, was the first to be identified. A reduction in ceruloplasmin (ferroxidase I) (EC 1.16.3.1) impairs the transport of iron to the erythropoietic sites, leading to hypochromic anemia. Hephaestin is located on the basolateral membrane of gut cells (Vulpe et al., 1999).The membrane bound form of ceruloplasmin is predominantly located in the brain (Jeong and David 2003),while zyklopen is central to trans-placental transfer of iron, though it is also located in retinal cells (Chen et al., 2010).

Deficiency of the cuproenzyme lysyl oxidase (EC 1.4.3.13) may be responsible for the skeletal and vascular defects reported in copper deficiency. Lysyl oxidase is essential for the normal maturation of collagen, specifically in the steps involving the formation of lysine-derived cross-links (O'Dell, 1976). The latter are required for the formation of strong, flexible connective tissues (Rucker and Murray, 1978).

Depigmentation is often associated with a deficiency of tyrosinase (EC 1.14.18.1), an enzyme involved in the production of melanin. The latter is necessary for protection against excess ultraviolet exposure. The central nervous system disturbances in copper deficiency are the result of myelinization derangements or of abnormal catecholamine levels and associated decreased activity of dopamine β monooxygenase (EC 1.14.17.1) or reduced activity of cytochrome c oxidase (EC 1.9.3.1) (Mason, 1979).

• Absorption and metabolism of copper

The major site of copper absorption is the small intestine, although some copper is also absorbed from the stomach. Absorption is inversely related to the amount of copper in the diet, decreasing the likelihood of deficiency and toxicity (Turnlund et al., 1989). For dietary intakes above 5mg/d, absorption is < 20%, whereas for intakes < 1mg/d, absorption is as high as 50% (Figure 24b.1).

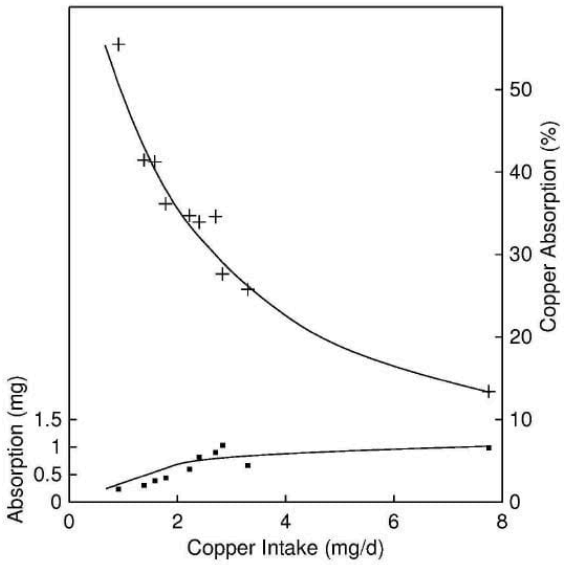


Figure 24b.1. Copper absorption as a percentage and in mg/d in relation to a dietary copper intake varying from 0.8–7.5 mg/d. Redrawn from Turnlund et al., American Journal of Clinical Nutrition 67: 960S–964S, 1998

As noted above, on the apical membrane, facing the gut, copper is taken up by a carrier mediated process, involving a transporter named Ctr1 (Kuo et al., 2006). Following uptake, the Cu is transferred across the cell and excreted into the portal circulation using an energy-dependent process and a Cu-ATPase, known as ATP7A (Harris, 2000). This enzyme is deficient in Menkes disease patients, and its absence is responsible for the subsequent copper deficiency. It is mostly located in the gut, but it also plays an important role in trans-placental transport and may be involved in uptake across the blood brain barrier (Hardman et al., 2004).

Once absorbed, copper is transported via the portal blood, bound to albumin and transcuprein, mainly to the liver. In the liver, copper is incorporated into ceruloplasmin, and then distributed to other tissues. For further details, see the review by Linder et al. (1998) and Harris (2000).

At present, the mechanisms of copper regulation are only beginning to be understood. Tissue copper concentrations are maintained within a narrow range, through homeostatic mechanisms that operate at the level of intestinal absorption and biliary excretion. Of the two levels, biliary excretion is the more important. As more copper is absorbed, turnover is faster, and, hence, more copper is excreted into the gastrointestinal tract. Very little copper is excreted in the urine, and urinary losses do not help regulate copper stores. Relatively little copper is stored in the body, compared with other trace elements such as iron (Turnlund, 1998; McArdle et al. 1999).

Copper is taken into the gut cell through Ctr1. In the cell, and in fact in all other cells apart from red blood cells, copper is bound to one of a series of chaperones. The first to be identified was Atox1, followed by several others, each of which takes copper from the membrane transporter to specific target proteins (Palumaa, 2013). The copper is transported across the cell, excreted into the portal circulation through the enzyme which is defective in Menkes disease, ATP7A and transferred to the liver bound to albumin, transcuprein or as a low-molecular weight complex with histidine or other amino acids (Gambling et al., 2011).

Within the liver, Cu is stored as metallothionein, re-excreted in the bile or incorporated into ceruloplasmin. Excretion and Cp incorporation both involve a second Cu-ATPase, ATP7B. ATPB is a membrane protein, which changes location in the liver ER, depending on whether the copper being transported is excreted or incorporated into ceruloplasmin (McArdle et al.,1999).

• Copper deficiency in humans

Copper deficiency in humans is rare. It has been described in infants recovering from malnutrition (Graham and Cordano, 1969) and in premature and low-birth-weight infants fed cow's milk (al-Rashid and Spangler, 1971; Griscom et al., 1971; Ashkenazi et al., 1973). Copper deficiency has also been reported in patients receiving prolonged total parenteral nutrition (TPN) unsupplemented with copper (Karpel and Peden, 1972; Dunlap et al., 1974; Vilter et al., 1974) and in a woman receiving a normal diet supplemented

with antacids (Anonymous, 1984). Severely handicapped patients (n=60) receiving an enteral diet with only 15µg/1000kcal Cu for 12–66mo also developed copper deficiency (Higuchi et al., 1988).

The earliest clinical manifestation of copper deficiency in humans is persistent neutropenia, usually followed by hypochromic microcytic anemia, scurvy-like bone changes, and osteoporosis (in infants) (Cordano et al., 1964; al-Rashid and Spangler, 1971; Karpel and Peden, 1972; Higuchi et al., 1988). The anemia probably results from defective iron mobilization and hemoglobin formation related to lower ceruloplasmin activity (Danks, 1988), or to increased susceptibility to oxidative damage (Johnson and Kramer, 1987).

Less frequent features of copper deficiency include hypopigmentation of the hair and hypotonia (low muscle tone), as well as impaired growth, an increased incidence of infections, and changes in the phagocytic capacity of the neutrophils (Uauy et al., 1998). In some experimental depletion-repletion studies, changes in blood clotting factors (Milne and Nielsen, 1996) and in markers of bone metabolism (Baker et al., 1999) have also been reported. Low copper intakes have been implicated as a risk factor in the development of cardiovascular disease (Lukaski et al., 1988; Klevay, 2000).

A few cases of adults with sickle cell disease who were receiving prolonged zinc therapy have developed hematological signs of copper deficiency(Prasad et al., 1978). Clinical features of copper deficiency, with the exception of anemia, occur in infants with Menkes kinky hair syndrome (KHS), an X-linked, recessive, hereditary disorder of copper metabolism with an intracellular defect of copper utilization (Danks et al., 1972).

Most of the reports of copper deficiency in humans have resulted from inadequate dietary intakes of copper, often associated with prolonged diarrhea, which prevents reabsorption of copper from the bile. Hence, it is not surprising that subjects with malabsorption syndromes — including tropical sprue, celiac disease, cystic fibrosis, and short-bowel syndrome resulting from intestinal resection — are at risk of copper deficiency (Rodriguez et al., 1985).

Diseases associated with chronic loss of proteins, such as nephritic syndrome and protein-losing enteropathy, may also cause copper deficiency as a result of loss of ceruloplasmin and of copper bound to albumin. High intakes of fructose and sucrose exacerbate the effects of copper deficiency, but the importance of this effect in humans is unclear (O'Dell, 1990).

• Food sources and dietary intakes

Copper is widely distributed in foods; the richest food sources are oysters, other shellfish, organ meats (e.g., liver, kidney), nuts, and dried legumes. Milk and dairy products are low in copper. The copper level in plants is not influenced by the copper content of the soil in which they grow (Institute of Medicine, 2001). Drinking water may contribute a relatively large amount of copper to the daily intake, especially if the water is slightly acidic or very soft and is supplied from copper pipes (Lönnerdal, 1996).

Copper deficiency among the general population has not been described. Dietary intakes of copper for adult European populations range from 0.9 to 2.3mg/d (van Dokkum, 1995). Data are available on the copper intakes of the U.S. population based on the Continuing Survey of Food Intakes by Individuals (CSFII 1994–1996) and NHANES III (IOM, 2001) and for the U.K. population, based on the National Diet and Nutrition Surveys (Gregory et al., 1990; Gregory et al., 1995; Gregory et al., 2000; Finch et al., 1998; Whitton et al., 2011). Copper intakes of infants fed breast milk or cow's milk are often low. More recently, the European Food Safety Authority has published data on dietary intake values in the human population (EFSA Panel on Dietetic Products 2015).

The composition of mixed western diets has little effect on the bioavailability of copper in humans. Copper deficiency has been observed in some cases of excessively high intakes of iron or zinc supplements (Fischer et al., 1984; Barclay et al., 1991; Botash et al., 1992) induced by their adverse effect on copper absorption. There is some evidence that the adverse effect of high dietary iron intake on copper absorption may occur only when copper status is low or marginal (Cohen et al., 1985).

• Effect of high intakes of copper

Under normal circumstances, copper levels in the body are tightly regulated. However, several cases of accidental copper toxicity in humans have been described, arising from ingestion of copper sulfate (Chuttani et al., 1965), acidic drinks in prolonged contact with copper (Paine, 1968), or drinking water with an unusually high copper concentration (8000µg/L) (Salmon and Wright, 1971). Adverse effects include nausea, gastrointestinal effects (nausea, diarrhea, cramps, abdominal pain), and, in some cases, liver damage (O'Donohue et al., 1993; Pizarro et al., 1999).

Wilson disease results from defects in ATP7B, the Cu-ATPase responsible for transferring copper from the liver to the bile or to the sites of ceruloplasmin synthesis. The defect results in copper overload in the liver and the brain. Patients suffering from Wilson's disease exhibit symptoms of copper toxicity such as nausea, vomiting, diarrhea, acute hemolytic anemia, hepatic necrosis, hepatic central vein dilation, and jaundice (Zelkowitz et al., 1980). In Wilson's disease, copper accumulates in the liver because the copper in the liver cells can neither enter the bile nor participate in ceruloplasmin synthesis (Linder and Hazegh-Azam, 1996; Suzuki and Gitlin 1999). See Czlonkowska et al., 2019 for a comprehensive review. Idiopathic copper toxicosis has been described in an Austrian population and termed Tyrolean copper toxicosis (Muller et al., 1998). It is hypothesized that the disorder arises as a combination of an inherited genetic mutation and high coper intake. A similar disease has been described by Tanner and colleagues in Indian children (Tanner 1998), although the situation is probably more complex than a simple copper overload (Pankit and Bhave 2002).There is no evidence that high dietary copper intakes are a public health problem. The U.S Tolerable Upper Intake Level (UL) for copper for adults > 19y is 10mg/d. For pregnant and lactating women and adolescents 14–18y, the UL is set at 8mg/d (IOM, 2001). Currently in Europe (2020), EFSA is examining upper levels of copper to harmonise permitted pesticide levels with nutritional guidelines.

• Indices of copper status

The methods used to assess copper status in humans include static tests, as well as the measurement of the activities of certain copper enzymes. These methods are discussed below. Other approaches use categorical regression to determine the optimal levels of copper in the diet (Chambers et al., 2010). There is still no consensus on the best biomarker to use, in part, because of the efficiency of the homeostatic control that maintains tissue copper concentrations within a narrow range (Harvey and McArdle 2008; Harvey et al., 2009; Danzeisen et al., 2007). As a result, measures of copper status tend to be resistant to change, except when dietary intakes are very low or very high. Serum copper or serum ceruloplasmin are most often used to assess copper status, but the sensitivity and specificity of both these measures are poor. Hence, they should be combined with a more sensitive and specific functional measure such as cytochrome c oxidase in platelets or superoxide dismutase in erythrocytes. It is likely that in the future functional indicators, perhaps including antioxidant status, changes in immune function, and the lysyl oxidase activity in skin, will be used in conjunction with other biochemical markers.

This page titled 27.1: Copper (24b.0) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.2: Serum copper (24b.1)

There are two main forms of copper in serum; one is firmly bound to ceruloplasmin, the other is reversibly bound to serum albumin. and to amino acids and other, poorly defined components (Linder 2016). Serum also contains the two copper enzymes cytochrome c oxidase and monoamine oxidase.

Serum copper concentrations are a reliable biomarker in cases of severe copper deficiency, when they are almost always low. However, they are not sensitive enough to be used as a measure of marginal copper status, as noted earlier, and do not reflect dietary intake, except when it is low. Normally, supplementation with copper does not increase serum copper because concentrations are under strict homeostatic control. Levels are only reduced after copper stores are significantly depleted. In healthy adults consuming diets adequate in copper, the lower end of the normal range for serum copper is 10µmol/L.

Experimental studies have also confirmed that serum copper is not a sensitive index of marginal copper status in adults (Reiser et al., 1985; Turnlund et al., 1989; Milne and Nielsen, 1996). In a depletion–repletion study of postmenopausal women fed a diet with only 0.57mg/d Cu for 105d, serum copper levels did not change significantly during the copper-depletion period (Figure 24b.2), despite the presence of other biochemical changes (Milne and Nielsen, 1996).

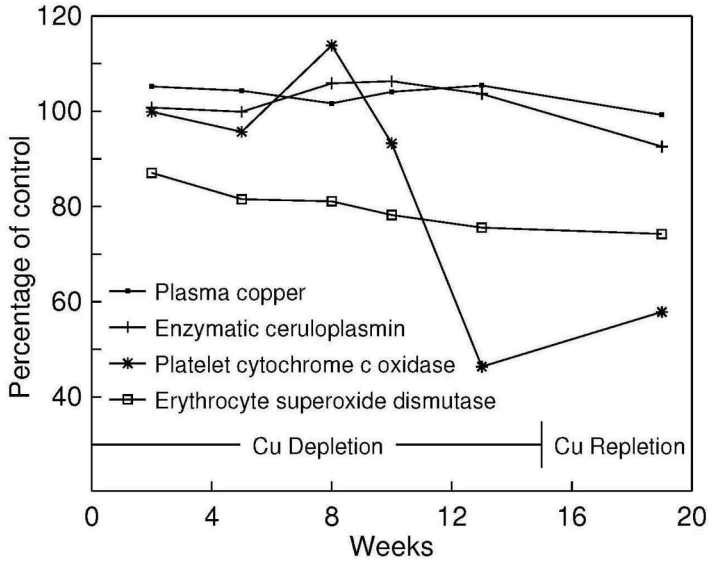


Figure 24b.2 Changes in copper biomarkers during copper deprivation. Each data point is the mean of 3wk surrounding date. Data from Milne and Nielsen, American Journal of Clinical Nutrition 63: 358–364, 1996

Eventually serum copper concentrations do fall. In patients receiving prolonged TPN unsupplemented with copper, falls of about 1.7µmol/L per week may occur (Solomons et al., 1976). Values < 9.5–11µmol/L have been reported in infants with copper deficiency (Cordano et al., 1964; Castillo-Duran and Uauy, 1988).Low serum copper concentrations also occur with the two genetic disorders of copper metabolism: Menkes kinky hair syndrome and Wilson's disease. In Menkes kinky hair syndrome, low serum copper concentrations arise from an intracellular defect of copper utilization, as discussed earlier (Danks et al., 1972). In Wilson's disease, the low serum copper levels are induced by defects in the hepatic storage of copper and in ceruloplasmin metabolism.

Plasma copper concentrations appear to be consistently lower than corresponding values in serum (Table 24b.2) (Smith et al., 1985). In contrast, concentrations of copper in serum and erythrocytes are comparable, so slight hemolysis does not affect serum copper concentrations.

Table 24b.2 Copper concentrations of serum or citrated plasma prepared from single venipuncture samples of eight subjects. Plasma was prepared by adding 0.15mL of 30% sodium citrate per 15mL of whole blood. Percentage difference = [(serum – plasma)/serum] × 100. Data from Smith et al., Journal of the American College of Nutrition 4:627–638, 1985

Subject no.	Serum (µmol/L)	Plasma (µmol/L)	Difference
1	20.0	19.4	3%
2	18.7	17.8	5%
3	19.0	17.3	9%
4	19.2	16.7	13%
5	19.5	16.7	15%
6	19.4	16.5	15%
7	18.1	16.5	9%
8	16.5	15.4	7%
Mean	18.9±1.1	17.0±1.1	9±4

More recently, examination of the non-ceruloplasmin pool of copper, the exchangeable fraction, has suggested that this might be a good indicator of copper overload, as in diseases such as Wilson's disease (Woimant et al., 2019).

• Factors affecting serum copper

It is important to remember that most assays do not differentiate between the different forms of copper in serum and usually, changes in “serum copper" are generally caused by alterations in the concentration of ceruloplasmin. The vast majority of coper in serum is bound to ceruloplasmin. Estimates vary from 60% to 95% with the remainder being bound to a poorly characterised series of chelators (Linder 2016). As well as being a ferroxidase, involved in iron metabolism, ceruloplasmin is an acute phase protein, and concentrations can be altered by other factors than copper status. This makes interpretation of the changes described below complex.

Age-related changes in serum copper concentrations are well documented. Newborn infants have low values, which rise to adult levels by 6 to 12mo of age (Salmenperä et al., 1986). In men, and in women not taking estrogen-containing preparations, serum copper levels increase with age until55–64y, after which they decline (Table 24b.3 (Fischer et al., 1990; Milne and Johnson, 1993).

Table 24b.3 Effect of age in men and women who were not taking estrogen-containing preparations on serum copper concentration and ceruloplasmin activity. Data from Fischer et al., Nutrition Research 10: 1081–1090, 1990.

Age group (years)	Men	Women
	Serum copper (µmol/L)	
25–34	14.7 ± 0.6	14.8 ± 0.5
35–44	15.0 ± 0.3	16.7 ± 0.4
45–54	16.2 ± 0.4	17.8 ± 0.6
55–64	16.3 ± 0.4	18.8 ± 0.7
65+	15.3 ± 1.0	16.8 ± 0.5
	Ceruloplasmin (U/dL)	
25–34	9.9 ± 0.7	11.0 ± 0.6
35–44	10.8 ± 0.3	12.8 ± 0.4
45–54	11.9 ± 0.4	15.2 ± 1.1
55–64	11.8 ± 0.5	14.1 ± 0.8

Age group (years)	Men	Women
65+	11.3 ± 1.2	12.3 ± 1.1

Gender influences serum copper. Levels in adult females tend to be higher than in males (Table 24b.3).

Pregnancy influences serum copper concentrations. Higher serum copper concentrations are evident after the third month of pregnancy (Halsted et al., 1968; Hambidge and Droegemueller, 1974). Whether this reflects changes in ceruloplasmin, which is part of the acute phase response, or an actual change in serum copper concentrations, is not always clear.

Estrogen-containing preparations, such as oral contraceptive agents and estrogen replacement therapy, elevate serum copper concentrations (hypercupremia) (Fischer et al., 1990; Nielsen et al., 1992). This effect is due to an estrogen-mediated rise in serum ceruloplasmin concentration No differences in copper absorption or copper balance have been reported between users and non-users of oral contraceptive agents (King et al., 1978; Crews et al., 1980). (Table 24b.4).

Table 24b.4 Effect of oral contraceptive agents (OCA) and estrogen replacement therapy on serum copper concentration and ceruloplasmin activity. Data from Fischer et al., Nutrition Research 10: 1081–1090,1990.

	Premenopausal women	
	Non-users of OCA	Users of OCA
Serum copper (µmol/L)	16.2±0.3	26.8±1.3
Ceruloplasmin (U/dL)	12.2±0.4	20.4±0.9
	Postmenopausal women	
	Non-users of estrogens	Users of estrogens
Serum copper (µmol/L)	18.1±0.4	25.9±1.2
Ceruloplasmin (U/dL)	14.5±0.7	19.7±1.0

Diurnal variation in serum copper concentrations has been noted (Cartwright, 1950; Guillard et al., 1979) with the highest levels occur in the morning (Lifschitz and Henkin, 1971; Cartwright et al., 1954; Yokoyama et al., 2000).

Regular strenuous exercise may also affect serum copper concentrations. Lukaski et al. (1983) showed values to be higher in male university athletes relative to controls.

Smoking raises serum copper concentrations (Kocyigit et al., 2001).

Infection, inflammation, and stress all elevate serum copper concentrations (Honkanen et al., 1991; Brown et al., 1993). This is a result of an increase in ceruloplasmin (an acute phase protein), effected by leukocytic endogenous mediators (Pekarek et al., 1972).

Malabsorption syndromes such as celiac disease, cystic fibrosis, and ulcerative colitis (Sternlieb and Janowitz, 1964) are often associated with low serum copper concentrations, arising from prolonged diarrhea which prevents reabsorption of copper from the bile (Rodriguez et al., 1985).

Certain disease states lead to alterations in serum copper concentrations. Levels increase in leukemia, Hodgkin's disease, various anemias, collagen disorders, hemochromatosis, and myocardial infarction and in patients with dilated cardiomyopathy (Mason, 1979; Oster, 1993). In other diseases (e.g., the nephrotic syndrome), serum copper concentrations are low (Cartwright et al., 1954).

• Interpretive criteria

Table 24b.5 Plasma copper concentrations in children, grouped according to age (mean±SD). Data from Rükgauer et al., Journal of Trace Elements in Medicine and Biology 11: 92–98, 1997.

Age group (years)	n	Serum Cu (µmol/L)
1 to < 2	15	21.1±4.6
2 to < 4	23	21.5±3.9
4 to < 6	19	19.4±5.3
6 to < 10	25	23.4±2.5
10 to < 14	21	21.1±3.7
14 to < 18	17	20.5±4.4

Interpretive criteria used for serum copper concentrations for adults are 8.8–17.5µmol/L for men and10.7–26.6µmol/L for women who are not taking oral contraceptive agents. A slightly higher range(15.7–31.5µmol/L) has been proposed for women taking estrogen-containing preparations (Milne and Johnson, 1993). Rükgauer et al. (1997) have published reference values for plasma copper for children ranging from 1mo to 18y, although the number of children studied in some of the age groups was small. These reference values are based on self-selected samples of apparently healthy subjects who were not taking any vitamin or mineral supplements (Table 24b.5). Salmenperä et al. (1986) have compiled reference values for serum copper in exclusively breastfed infants followed longitudinally for the first year of life. These reference values are shown in (Figure 24b.3). Serum copper levels were measured in subjects from 3 to 74y in the Canada Health Survey (Health and Welfare Canada, 1981). The National Health and Nutrition Examination Survey (NHANES II) also collected data on serum copper concentrations, but were not measured in the more recent U.K. national surveys.

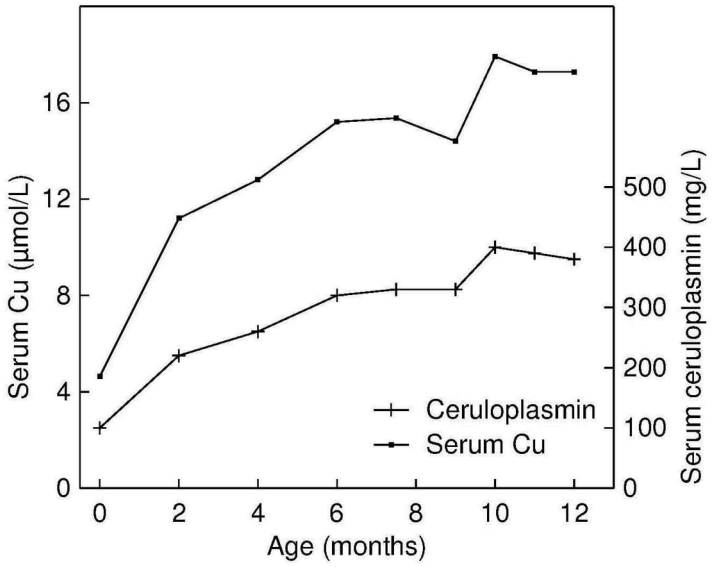


Figure 24b.3 Median serum copper and ceruloplasmin concentrations in exclusively breastfed infants during the first year of life. Data from Salmenperä et al., American Journal of Clinical Nutrition 43: 251–257, 1986.

• Measurement of serum copper

The most widely used method for measuring serum copper is AAS, although inductively coupled mass spectrometry is becoming more common (McMillen et al., 2009). Generally, a direct technique is used, involving sample dilution with deionized water (1 part plasma or serum to 1 part deionized water) (Osheim, 1983; Smith et al., 1985): a signal-enhancing mixture such as butanol and water is also sometimes used(Meret and Henkin, 1971). In some instances, use of a “high solids” burner head may be necessary

(Boling, 1966).Sometimes, the protein in the blood sample is removed using an acid such as trichloroacetic acid (Kelson and Shamberger, 1978),but this procedure may introduce volume errors during the deproteinization step and adventitious contamination from the acid and is not recommended. Analytical variation for serum copper by AAS is usually small; Gallagher et al. (1989) reported it to be < 2.6%. For very small pediatric samples, graphite furnace AAS, with Zeeman-effect background correction, may be used (Wang and Demshar, 1993). Increasingly inductively coupled plasma spectrometry (ICP) is being used because it is a multi-element technique with a detection limit and sensitivity better than flame AAS (Nixon et al., 1986). Standard reference materials for the analysis of copper in serum are available from NIST, Gaithersburg, MD.

This page titled 27.2: Serum copper (24b.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.3: Serum ceruloplasmin (24b.2)

Ceruloplasmin is the major copper-containing protein in the α2-globulin fraction of human serum. More than 60% to 95% of serum copper is associated with ceruloplasmin (Wirth and Linder, 1985), so changes in serum copper generally parallel the level of ceruloplasmin in the blood.

Ceruloplasmin (EC 1.16.3.1) is a copper transport protein synthesized primarily in the liver. It is a single-chain glycoprotein with six copper atoms per molecule. A ferroxidase, it assists in iron transport by oxidizing intracellular Fe^{2+} to Fe^{3+}, which can then combine with transferrin. Ceruloplasmin is also among those enzymes involved in the acute-phase reaction of inflammation and in the scavenging of oxygen radicals to protect cells against oxidative damage (Linder and Hazegh-Azam, 1996).

Ceruloplasmin levels in serum, like copper, are significantly reduced in cases of severe copper deficiency that can occur with the hereditary diseases associated with disturbances in copper metabolism: Menkes kinky hair syndrome and Wilson's disease. However, the response of ceruloplasmin to marginal copper deficiency or short-term copper depletion is variable (Milne and Nielsen, 1993; 1996). In most short-term studies of copper depletion (e.g., Milne and Nielsen, 1996) or supplementation (e.g., Kehoe et al., 2000), there has been little or no response by ceruloplasmin (Figure 24b.2).

Many of the factors that influence serum copper also affect serum ceruloplasmin and limit its usefulness as a marker of copper status, except in cases of severe copper deficiency. Furthermore, the complications and variations in measuring ceruloplasmin and copper in disorders such as Wilson's disease mean that their value as biomarkers need to be carefully assessed (Walshe, 2010).

• Factors affecting ceruloplasmin

Age and gender influence serum ceruloplasmin activity in the same manner as serum copper as shown in a Table 24b.3 (Fischer et al., 1990).

Estrogen-containing preparations tend to increase the serum ceruloplasmin activity, as shown in Table 24b.4 (Fischer et al., 1990). Ceruloplasmin synthesis and secretion by the liver is regulated in the long term by estrogen, probably through stabilization of mRNA (Middleton and Linder, 1993).

Pregnancy can increase serum ceruloplasmin activity (Fischer et al., 1990) again via the effect of estrogen on ceruloplasmin synthesis. This effect may be related to the copper needs of the growing fetus during pregnancy (Lee et al., 1993).

Infection and inflammatory stress can lead to markedly elevated ceruloplasmin activity levels in serum (Brown, 1998). Ceruloplasmin is an acute-phase protein and the inflammatory hormones (interleukin-1 [IL-1], IL-6, and tissue necrosis factor) regulate its acute synthesis and secretion by the liver (Linder and Hazegh-Azam, 1996). Hence, it is not surprising that serum ceruloplasmin activity is also elevated in rheumatoid arthritis (DiSilvestro et al.,1992).

• Interpretive criteria

Reference data for serum ceruloplasmin are limited. Salmenperä et al. (1986) have published reference values for exclusively breastfed infants followed during the first year of life, and they are shown in Figure 24b.3. Levels increase steadily during the first year of life, paralleling the trend for serum copper.

Milne (1994). suggests that a ratio of enzyme activity (ceruloplasmin oxidase activity) to ceruloplasmin protein concentration may be a useful index of copper status. The ratio has several advantages. Effects of age, sex, or hormone use are small (Table 24b.6)

Table 24b.6 Effect of age, gender, and hormone use on mean ceruloplasmin concentrations, measured both enzymatically (ENZ) and by radial immunodiffusion (RID). The ratio of ENZ/RID is also shown. FOC, females taking oral contraceptives; FE,females using estrogen therapy. Significance levels from analysis of variance (ANOVA) are also shown (interactions between gender and decade were nonsignificant in all three cases). Data from Milne and Johnson, Clinical Chemistry: 39, 883–887, 1993

Age (years)	Group	Enzyme activity (mg/L)	RID (mg/L)	ENZ/RID
20–29	M (n = 10) F (n = 10) FOC (n = 10)	408 503 612	266 330 371	1.56 1.53 1.71
30–39	M (n = 10) F (n = 10) FOC (n = 11)	416 516 744	296 314 525	1.49 1.66 1.42
40–49	M (n = 11) F (n = 11)	442 519	337 359	1.31 1.48
50–59	M (n = 9) F (n = 13) FE (n = 4)	441 516 593	310 370 384	1.43 1.44 1.57
60–69	M (n = 8) F (n = 10)	470 559	284 354	1.70 1.60
70+	M (n = 7) F (n = 7)	485 470	332 368	1.51 1.29
ANOVA	Gender Decade	0.0001 NS	0.0002 0.03	NS NS

(Milne and Johnson, 1993), and it appears to be more sensitive to changes in copper status than either the enzymatic activity or immunoreactive protein alone (Milne et al., 1988). Indeed, in some individuals, ceruloplasmin oxidase activity (but not ceruloplasmin protein concentration) was depressed during experimental copper depletion (Milne et al., 1988). Further, in young women, ceruloplasmin oxidase activity has been reported as inversely related to physiological changes due to copper depletion (e.g., autonomic cardiovascular function)(Lukaski et al., 1988).

Table 24b.6 presents mean ceruloplasmin values measured both enzymatically (ENZ) and by radial immunodiffusion, and the ratio of the enzyme activity to the immunoreactive protein for males and females by age and hormone use (Milne and Johnson, 1993). Values for ceruloplasmin for the two assays were standardized to mg/L by using purified human ceruloplasmin as a standard.

• Measurement of serum ceruloplasmin

Serum ceruloplasmin can be assayed enzymatically by measuring p-phenylenediamine oxidase activity(Sunderman and Nomoto, 1970). Ceruloplasmin protein levels can be measured immunochemically by radial immunodiffusion techniques (Mancini et al., 1965; Buffone et al., 1979). Alternatively, nephelometry(Buffone et al., 1979) or immunoelectophoresis techniques (Gibbs and Walshe, 1979) and ELISA (various commercial kits are available) can be used. Some discrepancies have been reported between radial immunodiffusion and nephelometry, which may be attributed to different sources of antibodies used for the assays (Buffone et al., 1979). Analytical variation for serum ceruloplasmin performed by radial immunodiffusion ranges up to 6% (Gallagher et al., 1989).

This page titled 27.3: Serum ceruloplasmin (24b.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.4: Erythrocyte superoxide dismutase (24b.3)

Approximately 60% of the copper found in erythrocytes is present in the cytosol as superoxide dismutase (Cu,Zn-SOD, EC 1.15.1.1). This enzyme, which contains both divalent copper and divalent zinc, has a molecular weight of approximately 32,500. It is important as a scavenger of the free radical O_2^- that causes damage to membranes and biological structures. It catalyzes the dismutation reaction, converting the highly reactive superoxide anion O_2^-, to less reactive hydrogen peroxide and oxygen:

$$O_2^- + O_2^- + 2H^+ \rightarrow SOD \rightarrow H_2O_2 + O_2$$

Table 24b.7 Effect of copper supplementation on mean (±SD) serum copper, ceruloplasmin, and erythrocyte superoxide dismutase in infants recovering from malnutrition. Data from Uauy et al., Journal of Nutrition 115:1650–1655, 1985.

Measure	Time	Study (n=8)	Control (n=9)
Plasma copper (µmol/L)	Before After	15.4±10.5 26.0±5.4	22.4±4.1
Ceruloplasmin (mg/L)	Before After	330±260 500±140	540±120
SOD (U/g Hb)	Before After	1073±312 1371±207	1461±451

Studies with both animal models and humans have shown that the activity of erythrocyte Cu,Zn-SOD is depressed in copper deficiency and that the activity is often, but not always, restored during copper repletion (Bettger et al., 1979; Okahata et al., 1980; Uauy et al., 1985; Milne et al., 1990; Milne and Nielsen, 1996). Table 24b.7 presents data on erythrocyte Cu,Zn-SOD activity in infants recovering from malnutrition and receiving marginal copper intakes: levels returned to normal after copper supplementation (Uauy et al., 1985) and there was a significant positive correlation between erythrocyte Cu,ZnSOD and plasma copper concentrations in these malnourished infants (Figure 24b.4). Such a correlation was not reported by Fischer et al., (1990). in their study of healthy adults, probably because the individuals studied had a copper status within the normal range.

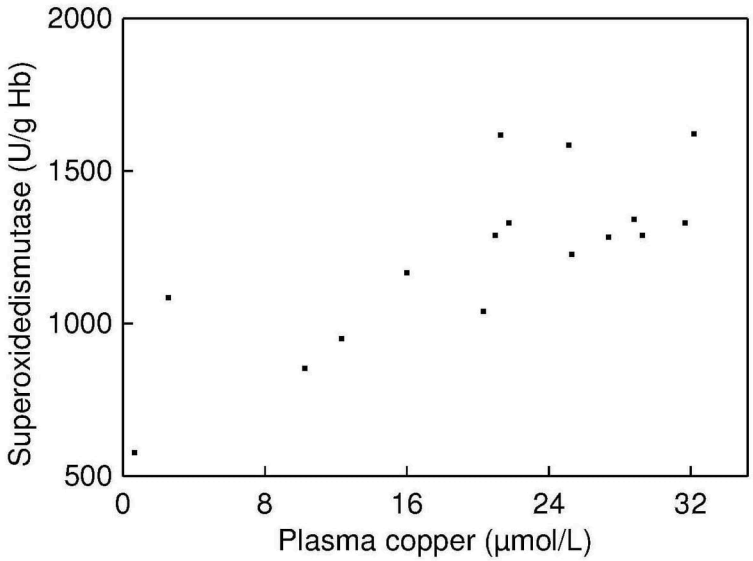


Figure 24b.4 Correlation between erythrocyte cell superoxide dismutase activity and plasma copper (r = 0.78, p<0.001) in eight malnourished infants studied before and after copper supplementation. Redrawn from Uauy et al., Journal of Nutrition 115: 1650–1655, 1985,

Positive correlations between erythrocyte Cu,Zn-SOD activity and other measures of copper status (e.g., liver cytochrome c oxidase) have also been observed in some animal studies (Bettger et al., 1979; Andrewartha and Caple, 1980).

Erythrocyte Cu,Zn-SOD activity appears to be a more sensitive index of copper depletion than serum copper or ceruloplasmin.

In several experimental copper depletion-repletion studies in humans (Reiser et al., 1985; Milne and Nielsen, 1996), erythrocyte Cu,ZnSOD activity declined significantly during the depletion phase, despite no detectable decrease in serum copper or ceruloplasmin. This trend is shown in Figure 24b.2.

Nevertheless, the response of erythrocyte Cu,Zn-SOD activity to copper repletion in these experimental studies has not been consistent. Several investigators have failed to show a response by erythrocyte Cu,Zn-SOD to copper supplementation (Turnlund et al., 1990; Medeiros et al., 1991). Even in a later study by Kehoe et al. (2000) involving two different copper compounds and two levels of copper supplementation, the activity of Cu,Zn-SOD in erythrocytes (and leukocytes) did not respond to the copper supplements.

It appears that the study duration, amount of copper fed, and probably rates of cell turnover all influence the response of erythrocyte Cu,Zn-SOD activity to copper repletion (Klevay et al., 1984; Milne et al., 1988; Turnlund et al., 1990; Medeiros et al., 1991; Milne and Nielsen, 1996; Kehoe et al., 2000). In Figure 24b.2 the erythrocyte Cu,Zn-SOD activity remained at a plateau during copper repletion, perhaps associated with the rate at which new erythrocytes were being synthesized. In contrast, in the study of malnourished copper-deficient infants (Table 24b.7), erythrocyte Cu,Zn-SOD activity increased to normal levels after supplementation with copper (80µg/kg/d) for 120d. Furthermore, Cu,Zn-SOD levels increase in a variety of diseases, including pancreatitis (Ściskalska et al. 2020). In some conditions that produce oxidative stress (Lukaski et al., 1990), including alcoholism and Down's syndrome (Del Villano et al., 1980; Porstmann et al., 1990), erythrocyte Cu,Zn-SOD activity may be elevated. In rheumatoid arthritis patients, lower levels than those in controls have been observed (DiSilvestro et al., 1992). Activity levels, but not total concentration, appears to be reduced in hyperglycemic, but not normal, diabetic patients possibly by glycosylation of the active site of the enzyme (Kotake et al., 1998).

Very few investigators have used erythrocyte Cu,Zn-SOD activity as an index of copper status in community-based studies. Fischer et al. (Fischer et al., 1990) measured erythrocyte Cu,Zn-SOD activity (and serum copper and ceruloplasmin) in 384 free-living adults and examined the relationships with age, sex, estrogen use, smoking, alcohol use, and exercise, as noted earlier. Table 24b.8 presents the mean ±SEM for erythrocyte Cu,Zn-SOD activity for males and females by age. Unlike serum copper levels and ceruloplasmin activity, erythrocyte Cu,Zn-SOD activity appears to be unaffected by age, gender (Table 24b.8), hormone use, or the acute phase status (Solomons, 1985; Fischer et al., 1990; Milne and Johnson, 1993).

Table 24b.8 Effect of age in men and women (not taking estrogen-containing preparations) on erythrocyte Cu,Zn-superoxide dismutase activity (U/mg Hb). The nonsignificant differences shown in the lower part of the table demonstrate the minimal impact of oral contraceptive use and estrogen replacement therapy on Cu,Zn-superoxide dismutase activity. Data from Fischer et al., Nutrition Research 10: 1081–1090, 1990.

Age group (years)	Men (Mean±SEM)	Women (Mean±SEM)
25–34	82.1±6.3	74.7±2.9
35–44	76.7±1.4	75.3±2.7
45–54	76.2±2.1	79.0±5.7
55–64	77.8±1.8	75.5±3.9
65+	77.4±2.9	68.9±3.9
Women Pre-menopausal	Non-users of oral contr. 75.1±2.1	Users of oral contr. 77.2±6.1
Women Post-menopausal	Non-users of estrogens 76.7±3.5	Users of estrogens 73.8±7.9

• Interpretive criteria

Interpretive criteria for erythrocyte Cu,ZnSOD activity are limited. Laboratories should establish their own reference range for erythrocyte Cu,Zn-SOD activity. Milne and Johnson (1993) have compiled ranges for healthy adults based on the assay of Winterbourne et al. (1975). However, there were relatively small numbers of subjects in each age group. Activity levels are unaffected by age, gender, and hormone use. The U.K. National Diet and Nutrition Survey of young people aged 4–18y measured

erythrocyte Cu,Zn-SOD activity; levels decreased for both boys and girls with increasing age. Mean, median, and the lower and upper 2.5 percentile values by age and sex are presented (Gregory et al., 2000). However, work is still required to confirm the utility of erythrocyte Cu,Zn-SOD activity in studies of the copper status of the general population.

• Measuring erythrocyte Cu,Zn-SOD

There is no standard assay for the measurement of erythrocyte Cu,Zn-SOD activity. Several methods are available; some are very time consuming, require large samples, necessitate the removal of hemoglobin immediately after sample collection, and are subject to interferences.

The assay generally involves the inhibition of oxidation-reduction reactions which are catalyzed by the superoxide anion. The latter may be generated enzymatically by, for example, xanthine plus xanthine oxidase, which, in turn, reduces cytochrome c. The reduction of cytochrome c is followed spectrophotometrically (Marklund and Marklund, 1974).

Several of these procedures are now automated. L'Abbé and Fischer (1986) developed a method for determining Cu,Zn-SOD activity that can be used with small samples. Commercial kits based on either an enzyme-linked immunosorbent assay or on a spectrophotometric assay are now available for measuring Cu,Zn-SOD activity in human serum and urine.

The analytical coefficient of variation for the erythrocyte superoxide dismutase assay when determined using an automated method ranges from 1.9% to 11.9%. The within subject variation for erythrocyte Cu,Zn-SOD activity is generally high (Gallagher et al., 1989).

This page titled 27.4: Erythrocyte superoxide dismutase (24b.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.5: Cytochrome c oxidase (24b.4)

Cytochrome c oxidase (EC 1.9.3.1) is present in most tissues of living organisms and is the terminal enzyme of the electron transport chain. It reacts directly with molecular oxygen in cellular respiration; it catalyzes the electron transfer from cytochrome c to O^2.

Cytochrome c oxidase activity in erythrocytes, platelets, or mononuclear leukocytes (lymphocytes and monocytes) has potential as a sensitive measure of copper status. Animal studies have generally reported reduced tissue cytochrome c oxidase activity as an early sign of copper deficiency(Underwood, 1971). in association with a reduction in the respiratory capacity of mitochondria, especially in the liver, heart, and brain. Coupling of respiration to phosphorylation is generally not affected (Linder and Hazegh-Azam, 1996). In some of these animal studies, cytochrome c oxidase activity in platelets correlated very strongly with liver copper levels ($r = 0.99$; $p < 0.0001$) (Johnson et al., 1993).

Genetic defects in the activity of cytochrome c oxidase, whereby activity is only about 50% of normal, may lead to neurological, cardiac, and muscle disease in children (DiMauro et al., 1985; Tulinius et al., 1991; Van Coster et al., 1991). Reduced activity of cytochrome c oxidase has also been reported in leukocytes in infants and children with Menkes kinky hair syndrome, a genetic defect in copper metabolism (Garnica et al., 1977).

More recently, results of copper depletion-repletion studies have reported that platelet cytochrome c oxidase activity may be more sensitive than either serum copper or ceruloplasmin to changes in copper intake. In Figure 24b.2 for example, the activity of cytochrome c oxidase in platelets was significantly lower ($p < 0.0001$) at the end of the copper-depletion period than at baseline Milne and Nielsen, 1996); levels for many women were below the reference range for comparable healthy women (Table 24b.9)

Table 24b.9 Effect of age, gender, and hormone use on mean cytochrome-c oxidase concentrations, measured in both platelets and mononuclear blood cells. FOC, females taking oral contraceptives; FE, females using estrogen therapy. Significance levels from analysis of variance (ANOVA) are also shown (interactions between gender and decade were nonsignificant in both cases). Data from Milne and Johnson, Clinical Chemistry 39: 883–887, 1993

Age (years)	Group	Platelets U/109 cells	Mononuclear blood cells U/106 cells
20–29	M F FOC	2.75 (10) 2.36 (9) 2.87 (9)	0.33 (9) 0.27 (7) 0.32 (7)
30–39	M F FOC	2.94 (9) 2.75 (9) 3.87(10)	0.31 (8) 0.36 (10) 0.39 (11)
40–49	M F	3.52 (9) 3.08 (10)	0.40 (10) 0.35 (9)
50–59	M F FE	3.89 (9) 3.88 (12) 3.54 (3)	0.35 (9) 0.38 (10) 0.42 (3)
60–69	M F	4.92 (7) 4.56 (9)	0.43 (6) 0.37 (9)
70+	M F	4.51 (6) 3.65 (5)	0.45 (4) 0.39 (7)
ANOVA	Gender Decade	NS 0.0001	NS 0.004

(Milne and Johnson, 1993). Moreover, unlike erythrocyte Zn,Cu-SOD activity or levels of serum copper and ceruloplasmin, the activity of platelet cytochrome c oxidase showed a positive response to copper repletion, with activity levels returning to about 60% of control values (Milne and Nielsen, 1996).

Cytochrome c oxidase activity responded similarly in both platelets and mononuclear leukocytes to copper repletion in young women fed a copper-depleted diet. However, baseline values were not measured, so that the response of these indices to depletion cannot be evaluated (Milne et al., 1988).

The activity of cytochrome c oxidase in both platelets and mononucleated leukocytes tends to be higher in older adults, and varies markedly between subjects (Table 24b.9, Milne and Johnson, 1993). The activity in mononuclear cells, but not platelets, was affected by oral contraceptive use, and only in the 30–39y age group. Adolescents with cystic fibrosis also have lower cytochrome c oxidase activity in mononucleated leukocytes, compared with age and sex matched controls (Percival et al., 1995). The postulated mechanism relates to the alterations in chloride transport in cystic fibrosis, which, in turn, impairs copper transport (Alda and Garay, 1990).

The interaction between copper metabolism and cytochrome c oxidase activity is complex and mediated by transporters, chaperones and affected by their metals and anions. For a recent review see Cobine et al. (2021).

• Interpretive criteria

Table 24b.9 presents mean cytochrome c oxidase activity in platelets and mononuclear blood cells (lymphocytes and monocytes) by age and sex for adults 20 to > 70y. Subjects were healthy, nonpregnant, and nonsmokers and were not taking any vitamin or mineral supplements, or prescribed medications, with the exception of estrogen.

• Measurement of cytochrome-c oxidase

Platelets and mononuclear cells have a shorter lifespan (≈10d) than erythrocytes (≈120d) and therefore activities in these shorter-lived cell types better reflect the metabolically active pool of endogenous copper. However, the isolation of such specific cell types is time consuming and complex, as noted in Section 15.1.3. Moreover, the enzyme cytochrome c oxidase is labile, and between subject variation is large. Hence, before the test can be used in community studies, the assay method must be rigorously standardized. An assay developed by Prohaska and Wells (1974) is often used, as are kits that measure ferrocytochrome to ferricytochrome colourimetrically.

This page titled 27.5: Cytochrome c oxidase (24b.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.6: Skin lysyl oxidase (24b.5)

Lysyl oxidase (EC 1.4.3.13) is another extracellular copper-containing enzyme that is essential for the normal maturation and function of collagen, specifically in the steps involving the formation of lysine-derived cross-links. The latter are required for the formation of strong, flexible connective tissue (Rucker and Murray, 1978).

Copper deficiency in animals results in low lysyl oxidase activity in heart, aorta, connective tissue, lung, and skin, which is reversed with copper supplementation (O'Dell, 1976). Decreased lysyl oxidase activity may be responsible for the skeletal and cardiac abnormalities reported in copper deficiency (Kagan and Trackman, 1991). In rats, despite no differences in bone weight or femoral ash content, torsional strength of the femur was markedly reduced in copper deficiency (Jonas et al., 1993).

To date, very few experimental copper depletion-repletion studies have examined the effect of low dietary copper on the activity of skin lysyl oxidase. Results in Figure 24b.5

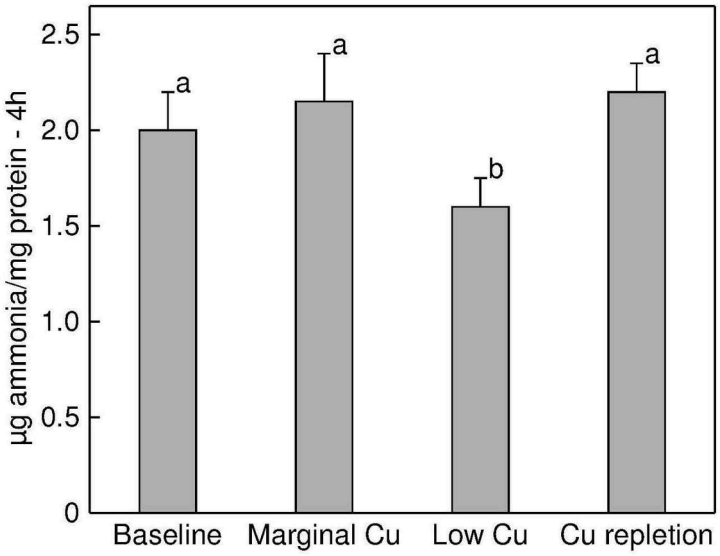


Figure 24b.5 Lysyl oxidase activity in the skin of healthy male subjects fed different amounts of copper. Marginal Cu, 0.66mg/d; Low Cu, 0.38mg/d; Cu repletion, 2.49mg/d. Bars with different letters differ significantly ($p<0.05$). Values are mean±SEM of 11 subjects at each time period. Redrawn from Werman et al., Nutritional Biochemistry 8: 201–204, 1997.

show the changes in mean lysyl oxidase activity in the skin of healthy males receiving diets containing four levels of dietary copper. There was a significant fall in activity when the subjects were fed a low copper diet (0.38mg/d), which returned to baseline values after copper repletion. No such decline was observed when the marginal copper diet (0.66mg/d) was fed. Plasma copper, ceruloplasmin, and urinary copper were also measured in this study; levels fell significantly on feeding the low-copper diet (Werman et al., 1997). The results of this initial study suggest that lysyl oxidase activity declines when dietary copper intake is inadequate, indicating that the cross-linking of collagen may be modulated by copper deficiency. Skin lysyl oxidase could be used as a biomarker of copper status in young men. Further studies are needed to assess whether the test can also be used in women and older subjects. Skin biopsies are invasive procedures, so that the test probably will have limited use in community-based studies.

This page titled 27.6: Skin lysyl oxidase (24b.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.7: Diamine oxidase (24b.6)

Diamine oxidase (EC 1.4.3.6) is a copper metalloenzyme that is responsible for the deamination of putrescine, cadaverine, and histamine (Wolvekamp and de Bruin, 1994). The activity of this enzyme has been reported to be low in some studies in rats (Kehoe et al., 2000) and renal dialysis patients with copper deficiency (DiSilvestro et al., 1997).

In an experimental copper supplementation study of healthy adults, the activity of diamine oxidase in serum increased significantly during all three Cu-supplementation regimens in both males and females, as shown in (Table 24b.10, Kehoe et al., 2000). These results suggest that the activity of diamine oxidase in serum has the potential to be a sensitive indicator of copper status in adults and warrants further investigation. Whether there are any adverse health consequences associated with low serum diamine oxidase activity is unknown.

Table 24b.10 Response of serum diamine oxidase (U/L) in men and women to three different regimens of copper supplementation. In all three cases and for both sexes, activity in the treatment group differed significantly from placebo (p<0.01). CuGC: copper gluconate. Abstracted from Kehoe et al., British Journal of Nutrition 84: 151–156,2000

Group (n)	Mean (SE)	Mean (SE)
Men (7) Women (9)	3mg $CuSO_4$/d 0.87 (0.27) 1.64 (0.47)	Placebo 0.27 (0.17) 0.33 (0.17)
Men (9) Women (11)	3mg CuGC/d 1.21 (0.36) 1.92 (0.41)	Placebo 0.47 (0.01) 0.42 (0.23)
Men (7) Women (10)	6mg CuGC/d 1.55 (0.49) 2.37 (0.79)	Placebo 0.44 (0.28) 1.16 (0.32)

The activity of this enzyme is, however, modified by conditions other than copper status, including pregnancy (Kusche et al., 1974), uremia (Tam et al., 1979), malabsorption syndromes or gastrointestinal damage(Bayless et al., 1981), and kidney dialysis (DiSilvestro et al., 1997).Further studies are needed with other population groups to establish the specificity of this enzyme to changes in copper status. A sensitive automated colorimetric assay developed by Takagi et al. (1994) can be used for the assay of serum diamine oxidase.

This page titled 27.7: Diamine oxidase (24b.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.8: Copper chaperones (24b.7)

Copper, once inside the cell, is bound to one of a series of chaperones, depending on its final target. The first to be identified was in yeast, and was termed ATX1 (Lin and Culotta 1995). Subsequently, several others have been identified and the best candidate for an indicator of status is CCS. In rat feeding studies, CCS in erythrocytes is responsive to both marginal and deficient diets (Bertinato et al., 2003) and also showed sensitivity to secondary copper deficeincy induced by Zn overload (Iskandar et al., 2005). Other labs confirmed these results in both rats and mice and in copper deficiency and excess (Prohaska et al., 2003). Prohaska and colleagues went one step further, suggesting the CCS:SOD1 ratio may be a more sensitive marker of status (Prohaska et al. 2003).

Araya and colleagues tested the hypothesis that CCS would not be affected by inflammatory processes, unlike ceruloplasmin, by measuring CCS mRNA and CCS protein in serum in volunteers who had inflammatory responses, as demonstrated by clinical history and C-reactive protein. They showed no change in either CCS protein or mRNA, supporting the idea that CCS could be used as a copper marker, and not as an indicator of ceruloplasmin (Araya et al., 2014).

This page titled 27.8: Copper chaperones (24b.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

27.9: Other indices of copper status (24b.8)

• Hair copper concentrations

Concentrations of copper in hair appear to correlate with levels in the liver (Jacob et al., 1978), heart, and kidney (Klevay, 1981). of rats. However, the validity of hair copper concentrations as a biomarker of copper status in humans is less certain. In a detailed human study of therosclerosis patients, hair copper concentrations did not correlate with the copper content of heart, muscle, liver, kidney, aorta, or rib (Aalbers and Houtman, 1985).

Many studies have confirmed that gender (in adults), pregnancy, lactation, prematurity, hair color, race, and age all affect hair copper concentrations (Petering et al., 1971; Creason et al., 1975; Gibson and DeWolfe, 1980; Taylor, 1986). The marked rise in hair (and serum) copper concentrations during early infancy is presumably associated with the redistribution of tissue copper which occurs at this time, and not with changes in dietary copper intake (Gibson and DeWolfe, 1980) (Figure 24b.6).

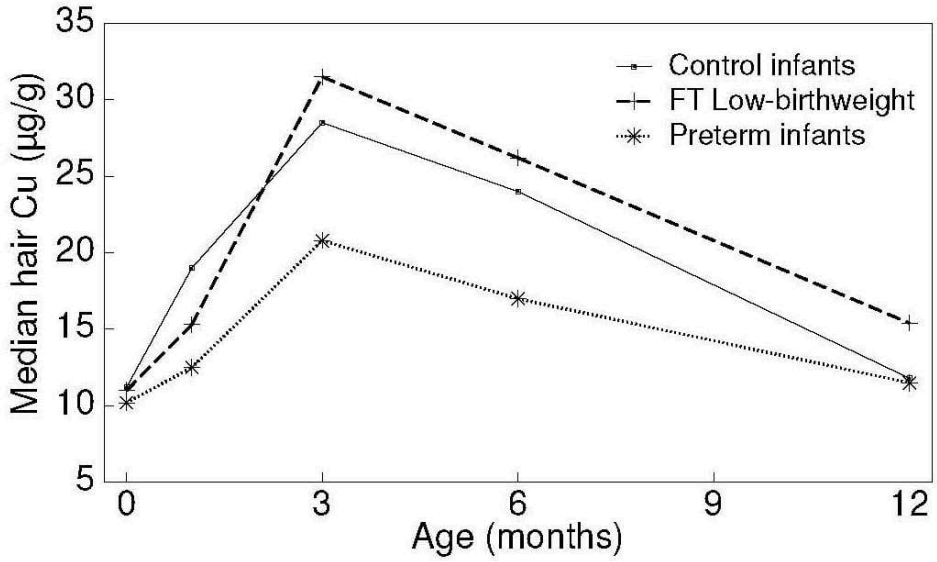


Figure 24b.6: Changes in hair copper concentrations during infancy. Redrawn from Gibson RS, (1982). The trace metal status of some Canadian full-term and low-birth-weight infants at one year of age. Journal of Radioanalytical Chemistry 70: 175–189,

Normal hair copper levels are found in two conditions in which copper accumulates in the liver — primary biliary cirrhosis (Epstein et al., 1980) and Wilson's disease (Rice and Goldstein, 1961; Gibbs and Walshe, 1965). Additionally, infants with copper deficiency, characterized clinically by neutropenia and biochemically by concentrations of serum copper and ceruloplasmin below the detection limit, did not have lower hair copper levels than age and sex-matched controls (Bradfield et al., 1980). Similarly, the copper content of the hair of children with Menkes kinky hair syndrome is normal (Danks, 1980). It appears, therefore, that hair copper cannot be used as a biomarker of copper status in humans.

• Urinary copper concentrations

Urinary copper is seldom used as an index of copper status; levels are very low in healthy subjects(10–60µg/d) because copper is efficiently reabsorbed by the renal tubules. The main excretory route for copper is the biliary system. Urinary copper concentrations do decrease in persons receiving TPN copper-deficient solutions (Solomons, 1979), conserving body copper. However, in a 13wk experimental depletion-repletion study in which men were fed three levels of dietary copper, urinary copper levels were unaffected by dietary copper levels (Turnlund et al., 1990).

• Other indices

Very limited data are available on copper concentrations in erythrocytes (Vitoux et al., 1999), and fingernails (Martin, 1964). Elevated values for erythrocyte copper have been reported in patients with the two inherited metabolic disorders associated with defects in copper metabolism (Wilson's disease and Menkes syndrome), as well as in children with Down's syndrome and cystic fibrosis. Further studies are needed to ascertain the validity of these measures of copper status.

Changes in some biochemical markers of bone metabolism have been noted during experimental copper depletion-repletion (Baker et al. 1999). and also in some supplementation trials (Baker et al., 1999; Cashman et al., 2001). Results suggest that changes in both urinary pyridinoline and deoxypyridinoline (biomarkers of bone resorption) may serve as additional functional markers of copper status in populations that consume diets low in copper; details of these biomarkers of bone resorption are given in Section 23.1.5.

Peptidylglycine α-amidating monooxygenase (PAM) is a cuproenzyme required for neuropeptide synthesis. Prohaska and colleagues have suggested its enzyme activity, but not protein levels, may be modulated by mild copper deficiency (Bousquet-Moore et al. 2010; Prohaska et al. 1995). .

• Multiple indices

Serum copper is frequently used to assess copper status in both clinical and community studies, but it is not very sensitive or specific (Harvey and McArdle 2008). Increasingly, because of its higher sensitivity and specificity, the activity of Cu,Zn-SOD in erythrocytes is used in combination with serum copper. In the future, the activity of cytochrome c oxidase in platelets or diamine oxidase in serum may be used in combination with serum copper, instead of erythrocyte Cu,Zn-SOD, if their validity as biomarkers of copper status in a community setting is confirmed.

This page titled 27.9: Other indices of copper status (24b.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28: ZINC (CHAPTER 24C)

Abstract

Multiple physiological functions are affected by zinc deficiency because of the involvement of zinc in numerous metabolic processes as a catalytic, structural, and regulatory ion. As a consequence, in zinc deficiency numerous nonspecific changes occur, including impairments in growth, immune function, appetite, taste acuity, and possibly neurobehavioural function. Zinc homeostasis in the body is tightly controlled over a wide range of zinc intakes, primarily by adjusting losses of endogenous fecal zinc. Such strict homeostatic control leads to difficulties in the assessment of zinc insufficiency and excess. The BOND Expert Zinc Panel recommended three indicators to identify populations at elevated risk of zinc deficiency — the prevalence of usual zinc intakes below the estimated average requirement, the percentage with low plasma / serum zinc concentrations, and the percentage of children < 5y who are stunted (i.e., HAZ < –2SD).Reference limits for plasma / serum zinc (depending on age, sex, time of day, fasting status) are available to identify populations at risk of zinc deficiency, provided the correct protocols are followed for the collection, separation, and analysis of plasma zinc. Moreover, in view of the poor specificity of impaired linear growth, assessment of this functional indicator should be made in combination with at least one of the other indicators. Assessment of marginal zinc deficiency in individuals is more difficult because of the absence of frank clinical signs and the lack of a reliable sensitive and specific biomarker. The recommended approach combines a medical history, qualitative dietary and clinical assessment, with plasma / serum zinc providing supplementary data. The BOND Expert Zinc Panel classified other potential zinc biomarkers into three groups: (1) potentially useful (hair zinc, urinary zinc, and neuro-behavioral function); (2) emerging biomarkers (nail zinc, zinc-dependent proteins, oxidative stress and DNA integrity; zinc kinetics, taste acuity), and (3) not useful (zinc-dependent enzymes, erythrocyte and leukocyte zinc); strengths and weaknesses of each are discussed.

28.1: Zinc (24c.1)

28.2: Dietary Zinc Intake (24c.2)

28.3: Serum zinc (24c.3)

28.4: Linear growth (24c.4)

28.5: Urine zinc (24c.5)

28.6: Hair zinc concentrations (24c.6)

28.7: Nail zinc (24c.7)

28.8: Metallothionein concentrations in circulating blood cell types (24c.8)

28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)

28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)

28.11: Oxidative stress and DNA integrity (24c.11)

28.12: Kinetic markers: pool sizes and turnover rates (24c.12)

28.13: Taste acuity (24c.13)

28.14: Zinc-dependent enzymes (24c.14)

28.15: Erythrocyte zinc (24c.15)

28.16: Leukocyte zinc (24c.16)

28.17: Salivary zinc (24c.17)

28.18: Multiple biomarkers (24c.18)

This page titled 28: Zinc (Chapter 24c) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.1: Zinc (24c.1)

In an adult 70kg male there is approximately 1.5–2.5g of zinc, of which over 80% is found in the skeletal muscle and bone. Much smaller amounts are present in the liver, gastrointestinal tract, skin, kidney, lung, prostate, and other organs, as shown in (Table 24c.1).

Table 24c.1 Zinc content of major organs and tissues in adult (70 kg) man. Adapted from Iyengar (1998).

Tissue	Zn conc. (mg/kg wet wt)	Total content (mg)	Percent of total body zinc
Skeletal muscle	50	1400	63
Skeleton			
—Bone	90	450	20
—Marrow	20	60	3
—Periarticular tissue	11	11	< 1
—Cartilage	34	30	1
Liver	40	72	3
Lung	40	40	2
Skin	15	39	2
Whole blood	6	33	1
Kidney	50	15	1
Brain	10	14	1
Teeth	250	11.5	1
Hair	200	4	< 1
Spleen	20	3.6	< 1
Lymph nodes	14	3.5	< 1
Gastrointestinal tract	15	1.8	< 1
Prostate	100	1.6	< 1
Other organs/tissues	Variable	50	2
Total		2240	100

Unlike other trace elements such as iron and copper, or minerals such as calcium, there are no large, sequestered, readily mobilizable stores of zinc that can be rapidly released in response to variations in dietary intakes of zinc. Instead, there is a small pool of rapidly exchangeable zinc which is located in bone, liver, pancreas, kidney, spleen and plasma and accounts for about 10% of whole-body zinc (King et al., 2000). This pool turns over completely about five times each day to provide zinc for zinc-dependent metabolic functions throughout the body. It is the loss of a critical but small amount of zinc from this pool that leads to the biochemical and clinical signs of zinc deficiency (King, 2018). Bone may also provide a passive reserve of zinc, especially during growth, so that when dietary zinc intakes are low, less of the zinc released during normal bone turnover, is re-deposited in the skeleton (Zhou et al., 1993). There is also some evidence that hepatic zinc, bound to metallothionein and accrued during gestation, can also be mobilized during early infancy to supplement the infant's needs for zinc (Zlotkin and Cherian, 1988; Coni et al., 1996).

24c.1.1 Functions of zinc at the cellular level

At the cellular level, zinc has three major metabolic functions — catalytic, structural, and regulatory (King and Cousins, 2014). More than 300 metalloenzymes require zinc as a catalyst for function or regulation. During zinc deficiency, their activity is decreased, although their structure does not change, and with the addition of zinc, enzyme activity is restored. Zinc metalloenzymes are involved in nucleic acid metabolism and cellular proliferation, differentiation, and growth; examples are listed in (Table 24c.2).

So far, no consistent changes in the activity of any zinc metalloenzymes have been linked with signs and symptoms of zinc deficiency in experimental zinc depletion-repletion studies in humans (King et al., 2015).

Table 24c.2: Selected zinc metalloenzymes.

Zinc metalloenzymes
Alcohol dehydrogenase
Alkaline phosphatase
Carbonic anhydrase
Carboxypeptidase
Deoxynucleotidyl transferase
DNA polymerase
Glutamic acid dehydrogenase
Malic acid dehydrogenase
Nucleoside phosphorylase
RNA polymerase
Tyrosine kinase
Zinc-copper superoxide dismutase

Zinc also plays an essential structural role in the folding of some proteins and a lack of zinc results in protein misfolding and the loss of function. A finger-like structure, known as a zinc finger motif, stabilizes the structure of a variety of proteins, including those involved in cellular differentiation or proliferation, signal transduction, cellular adhesion, or transcription. Zinc fingers use cysteine and histidine to form a tetrahedral Zn^{2+} coordination complex. The extent to which the intake of dietary zinc affects the function of zinc finger proteins is unclear. Zinc also maintains the structure of certain enzymes, notably the antioxidant enzyme, copper-zinc superoxide dismutase (Cu,Zn-SOD) (King, 2011).

Over 3% of all identified human genes contain zinc finger proteins that regulate gene expression by acting as transcription factors. Gene expression may also regulate the effects of zinc deficiency on lipid peroxidation, immune function, apoptosis, and neuronal function(King et al., 2015). Zinc is also involved in cell signaling through the metal response element-binding transcription factor 1 (MTF1). Whether MTF1 regulates genes negatively or positively appears to depend on the cellular zinc status. Zinc may also have a direct regulatory role controlling numerous cell-signaling pathways by modulating kinase and phosphorylase activities. For more details of the metabolic functions of zinc at the cellular level, see King and Cousins (2014).

Zinc-dependent cellular functions are coordinated by one or more of twenty four zinc transporters. Two classes of zinc transporters exist: ZIP proteins and ZnT proteins. It is the up- or down-regulation of these transporter genes in response to zinc intake that helps to maintain cellular zinc homeostasis. The ZnT proteins (n=10) generally transport zinc ions out of the cytosol, whereas ZIP (SLC 39) proteins (n=14) import the ions from the cellular compartments or the extracellular space into the cytosol. Individual ZIP and ZnT transporters are located in specific cell types. For example, ZIP-14 appears to be involved in the uptake of zinc by the liver in response to acute inflammation and infection, whereas ZnT3 is found in the synaptic vesicles of some types of neurons, and ZnT8 in the secretory granules of β-pancreatic cells (King et al., 2015). However, the role of many of the twenty four zinc transporters and how their expression is altered to maintain cellular zinc homeostasis remains uncertain (King and Cousins, 2014).

24c.1.2 Physiological functions of zinc

Multiple physiological functions are affected by zinc deficiency due to the ubiquitous involvement of zinc in cellular metabolism. A modest deprivation in zinc intake over the long-term results in detectable changes in growth, immune function, and possibly neurobehavioural function. Other signs of marginal zinc deficiency include poor appetite and impaired taste acuity (hypogeusia).

Impaired growth in zinc deficiency arises from the role of zinc in DNA transcription and gene expression, signal transduction pathways, and endocrine function (e.g., insulin-like growth factor); these effects are summarized in Figure 24c.1.

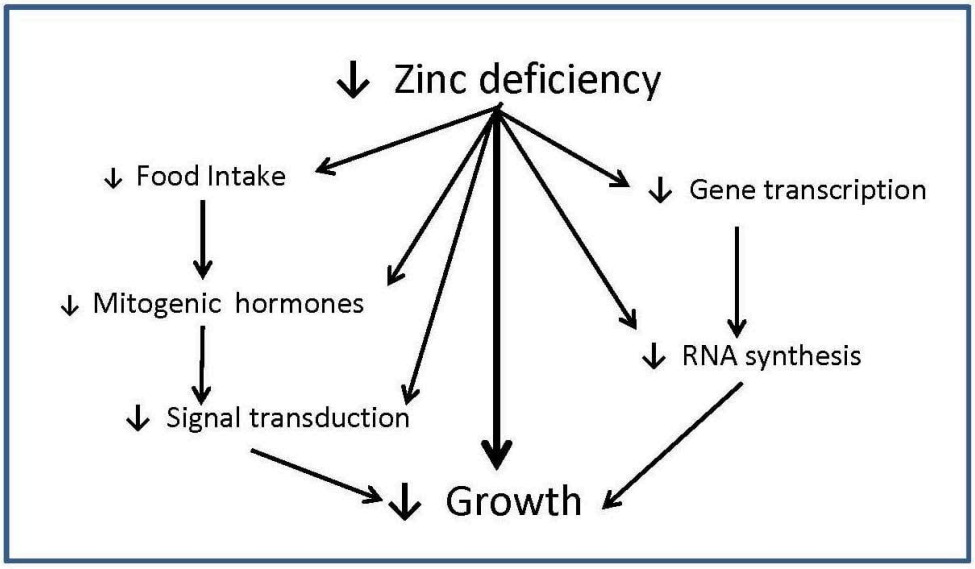


Figure 24c.1. Effects of zinc deficiency on metabolic processes associated with growth. Redrawn from MacDonald RS (2000).

Of the growth indices, height- or length-for-age is considered the best functional bioindicator associated with the risk of zinc deficiency in populations. Increased linear growth is the primary response to increased zinc intake, with an associated gain in weight. (de Benoist et al., 2007).

Zinc is also required for the synthesis and functions of immune regulatory proteins and for the maintenance of normal immune function. Consequently, zinc deficiency causes immune dysfunction. Both barrier function and cellular components of the innate immune responses (i.e., macrophage and neutrophil function) are compromised as well as the acquired immune system, with the largest effect being on reducing CD4 T cell function (King et al., 2015). The decline in innate immunity with zinc deficiency promotes systemic inflammation, characterized by elevated levels of proinflammatory cytokines. Supplementation with zinc has been associated with decreased inflammatory responses in children with zinc deficiency(Sandstead et al., 2008). With a reduced resistance to infection, the rate and duration of infections increases. Zinc deficiency also increases oxidative stress and DNA damage which, in turn, play a role in promoting the inflammatory process (Wong and Ho, 2012). For more details of zinc and immune function, see Raiten et al. (2015).

The mechanism whereby zinc influences neurobehavioral functioning is not well established. In the central nervous system, zinc is concentrated in the presynaptic vesicles of zinc-containing neurons, which are found primarily in the forebrain and connect with other cerebral cortices and limbic structures. During synaptic events, zinc is released and passes into postsynaptic neurons, serving as neurotransmitters. Zinc deficiency may interfere with these processes and thus compromise neurobehavioral functioning, especially during times of rapid growth and development, such as infancy, when demands for zinc are high (Black, 2003). Meta-analysis, however, has failed to show a significant overall effect of zinc intake on biomarkers of cognitive function in children, although the number of available studies in this review was small (n=6) (Warthon-Medina et al., 2015). There is some evidence that zinc deficiency may also be related to anxiety and depression. Possible mechanisms that warrant investigation include the involvement of zinc in the uptake of the neurotransmitter serotonin (5-HT) in the brain (Levenson, 2006). More research is also needed to clarify the mechanisms by which zinc affects taste acuity (Aliani et al, 2013). Some investigations suggest that reduced taste acuity may be linked with reductions in certain zinc metalloproteins that act as olfactory receptors (King and Cousins, 2014).

24c.1.3 Absorption and metabolism of zinc

Zinc is mainly absorbed in the upper part of the small intestine by two mechanisms: a saturable, carrier-mediated process and, secondly, by a non-mediated, or passive process. Two families of zinc transporters control the transport of zinc on the enterocyte membrane and account for the saturable component. Once zinc is absorbed into the enterocyte, movement of zinc within the cell is influenced by metallothionein and the transporter ZnT7. Export of zinc from the intestinal epithelial cells into the portal vein

is controlled by the ZnT1 transporter. About 30%–40% of the zinc in portal blood is exchanged with the liver, the main organ involved in zinc metabolism. From the liver, zinc is released into the systemic circulation for delivery to peripheral tissues bound mainly to albumin and α2-macroglobulin (King and Cousins, 2014). Almost no zinc circulates in an unbound state.

Absorption of zinc by a nonsaturable or passive mechanism only occurs at high intestinal luminal zinc concentrations generated through consumption of pharmacologic doses of zinc (King and Cousins, 2014). At such high doses (i.e., > 20mg/d Zn), fractional absorption of zinc via the passive mechanism is very low, suggesting that a dose of supplemental zinc above 20mg will have a limited influence on zinc nutrition (Tran et al., 2004).

Zinc homeostasis in the body is tightly controlled over a wide range of zinc intakes. This is achieved primarily by adjusting zinc excretion through the gastrointestinal tract. Losses of zinc via the gastrointestinal tract comprise both unabsorbed dietary zinc and endogenous fecal zinc. The sources of endogenous fecal zinc are uncertain but probably include secretions from pancreatic and intestinal mucosal cells. When zinc intakes are very low, there is a decrease in zinc secretion into the intestinal lumen, and endogenous fecal zinc declines. With increasing zinc intakes, endogenous fecal zinc losses increase to maintain homeostasis. If the adult zinc intake is extremely low (e.g., < 2mg/day), then the reductions in losses of endogenous fecal zinc may be insufficient to re-establish zinc homeostasis. In these circumstances, additional metabolic adjustments occur to mobilize zinc from the small, exchangeable pool located in liver, pancreas, kidney, spleen, and possibly bone, to maintain zinc-dependent functions (King, 2011).

Unlike iron, absorption of zinc does not change in response to alterations in whole-body zinc homeostasis or status. Instead, zinc absorption is influenced by current zinc intake, not the past or long-term zinc intakes, or status (Chung et al., 2008; King, 2010). Figure 24c.2.

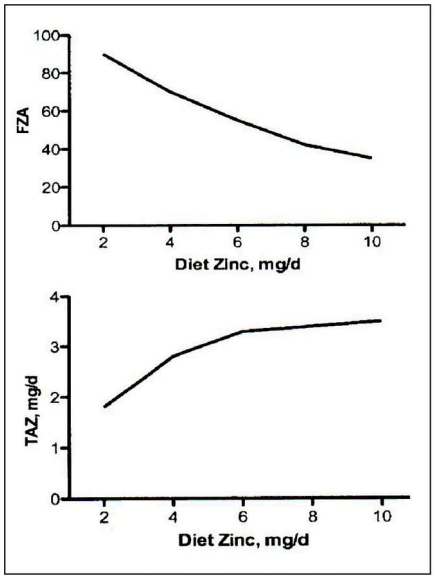


Figure 24c.2 Effect of dietary zinc on fractional (FZA) and total zinc absorption (TAZ}. Redrawn from King (2010).

Absorption of zinc can be measured / defined in two ways: (1) the fraction (or percent) of zinc ingested that is absorbed (Fractional Zinc Absorption or FZA), determined by the dividing the amount absorbed by the amount digested, and (2) the total quantity of absorbed zinc (TAZ) over the whole day. The fraction of zinc absorbed depends on the bioavailability of the ingested zinc, whereas the total quantity absorbed is dependent mainly on the amount of zinc consumed and the bioavailability of the zinc in the diet. Intake of dietary zinc affects fractional zinc absorption and the total quantity of absorbed zinc differently. With increasing zinc intakes, the total amount of absorbed zinc increases while the percent absorbed declines (King, 2010, Figure 24c.2). These adjustments in the efficiency of zinc absorption with changes in zinc intake are controlled by the up-regulation and down-regulation of zinc transporters and possibly other proteins involved in the transport of zinc (King and Cousins, 2014; King et al., 2015).

In addition to the amount of the dietary zinc intake, the food sources of zinc may also affect zinc absorption. The dietary factor with the greatest effect on zinc absorption is phytic acid (myoinositol hexakisphosphoric acid (InsP6) and its associated magnesium, potassium, and calcium salts — termed phytates. Phytic acid and its salts are found in high concentrations in

unrefined cereals, legumes, nuts and oil seeds. Phytate binds zinc in the intestinal lumen and forms an insoluble complex that cannot be digested or absorbed by humans because of the absence of the intestinal enzyme phytase (Iqbal et al., 1994). This inhibitory effect on zinc absorption can be substantial. If the habitual diet is rich in phytate, adults cannot adapt to increase zinc absorption (Hunt et al., 2008) or to enhance reabsorption of endogenous zinc (Hambidge et al., 2010). Whether phytate has an inhibitory effect on zinc absorption in young children is uncertain. Miller et al. (2015) failed to detect a negative effect of phytate on zinc absorption in their isotope studies of infants and young children. In contrast to phytate, the effect of dietary intakes of protein, calcium, iron and organic acids on zinc absorption appears limited or non-existent (IZiNCG, 2004; Miller et al., 2013).

Absorption and/or utilization of zinc may also be influenced by physiological, host-related, and contextual factors over and above that of dietary factors (Krebs et al., 2014), Figure 24c.3. Studies have shown an increase in zinc absorption in late pregnancy and lactation, most notably when dietary zinc intake is low (Fung et al., 1997; Donangelo et al., 2005; Donangelo and King, 2012), and possibly a decline with aging (Turnlund et al., 1986).

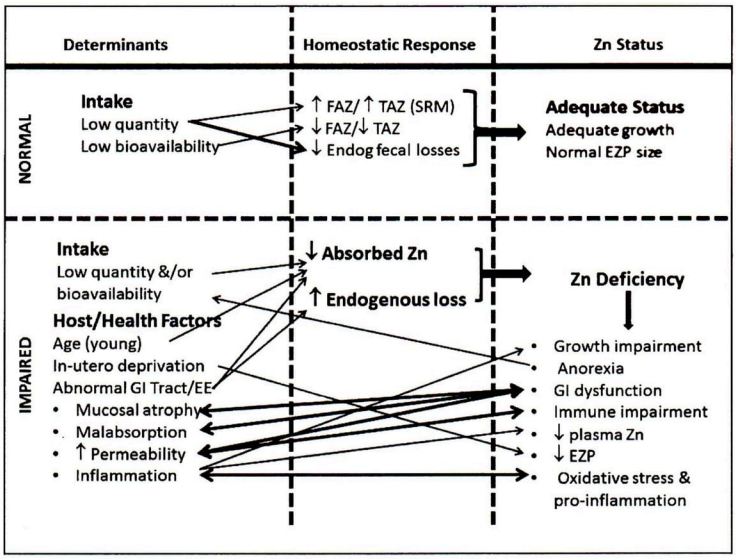


Figure 24c.3. Conceptual diagram to illustrate determinants of zinc status, the gastro-intestinal homeostatic response, and the features associated with normal and abnormal zinc status. In the condition of impaired zinc status, the thick arrows reflect bidirectional, co-existing efects between determinants and zinc deficiency. FAZ: fractional absorption of zinc; TAZ: total absorbed zinc; SRM: saturation response model; EE: environmental enteropathy; EZP: exchangeable zinc pool. From(Krebs et al., 2014).

Not surprisingly, given the dominant role of the gastrointestinal tract in zinc homeostasis, malabsorptive disorders that alter the integrity of the mucosal cells, reduce zinc absorption, although the magnitude of their effect is uncertain. An exception are patients with celiac disease in whom impairments in absorption and increases in intestinal endogenous zinc losses have been quantified through stable isotope studies (Crofton et al., 1990; Tran et al., 2011). Accumulating evidence suggests that these same disturbances in zinc absorption occur in environmental enteric dysfunction (EED), a chronic inflammatory condition linked to chronic, subclinical exposure to fecal pathogens (Syed et al., 2016). However, again, the magnitude of the effect of EED has yet to be quantified (Manary et al., 2010; Lindenmayer et al., 2014).

24c.1.4 Zinc deficiency in humans

When homeostatic mechanisms fail to ensure that requirements for zinc are met, clinical symptoms of zinc deficiency ensue. The first cases of dietary zinc deficiency in humans were described in the 1960s in male dwarfs from the Middle East (Prasad et al., 1963). Typical clinical features, which were corrected by zinc supplementation, included growth retardation, delayed secondary sexual maturation (hypogonadism), poor appetite, mental lethargy, and skin changes. In North America and New Zealand, overt and severe nutritional zinc deficiency was first recognized in hospital patients who received either parenteral nutrition or enteral feedings without zinc supplements (Kay and Tasman-Jones, 1975; Arakawa et al., 1976).

Two genetic disorders are known to induce zinc deficiency: acrodermatitis enteropathica and sickle cell disease (King et al.,

2015).

Secondary zinc deficiency has been documented in the presence of Crohn disease, inflammatory bowel disease, ulcerative colitis, and malabsorption syndromes. A variety of other diseases — including renal and liver diseases, diabetes, acquired immuno-deficiency syndrome, and alcoholism — also induce secondary zinc deficiency. In such cases, zinc deficiency may arise from either increased urinary excretion of zinc (hyperzincuria) or excess endogenous fecal zinc losses (Aggett and Comerford, 1995).

Treatment with certain drugs has also been associated with secondary zinc deficiency. For example, long-term treatment with penicillamine for Wilson's disease (Smolarek and Stremmel, 1999) diethylenetriamine pentaacetate for the treatment of iron overload in thalassemia and sickle cell anemia (King and Cousins, 2014), and the administration of chlorthiazide (used to treat edema) and glucagon (used to control blood glucose levels) (Prasad and Oberleas, 1976) have all been implicated as iatrogenic causes of zinc deficiency.

In some industrialized countries marginal zinc deficiency, characterized by slow physical growth, has been identified in apparently healthy infants and children (Hambidge et al., 1972a; Walravens and Hambidge, 1976; Walravens et al., 1983; Smit-Vanderkooy and Gibson, 1987; Gibson et al., 1989a). More recent double-blind preventative zinc supplementation studies in both industrialized and low-income countries have confirmed an increasing range of other functional impairments associated with zinc deficiency in children, in addition to impaired growth (Brown et al., 2009). In many of these studies, abnormalities of the immune system and increased risk of some common childhood infections have been reported. In settings where zinc deficiency is likely to be a problem, reductions in the incidence of diarrhea following preventive zinc supplementation have been observed, although the impact on lower respiratory infections, including pneumonia, has been less consistent (Haider and Bhutta, 2009; Brown et al., 2009; Lassi et al., 2010; Roth et al., 2010; Mayo-Wilson et al., 2014; Liu et al., 2018), and no clear effect on the incidence malaria (Brown et al., 2009; Mayo-Wilson et al., 2014). Therapeutic zinc supplementation has been shown to reduce the duration of acute and persistent diarrhea (Bhutta et al., 2000; Haider and Bhutta, 2009) , and is now recommended by World Health Organization to be included in diarrhea control programs (WHO/UNICF, 2004).

Reports of zinc-related neurobehavioral abnormalities in children and/or adults include delays in child development, anorexia, dysfunction of smell and taste, irritability, mood and depression. However, studies of zinc deficiency or zinc supplementation on neurobehavioral function, have produced inconsistent results, possibly due in part to the differences in the measurement methods used (Warthon-Medina et al., 2015). In a meta-analysis of trials in infants and young children, for example, analysis of a small subset of randomized controlled trials (n=6) showed no significant effect of zinc on cognitive function or motor skills (Warthon-Medina et al., 2015). In contrast, results of a randomized controlled trial on Peruvian infants suggested that supplemental zinc may support normative neuro-development in infants consuming low zinc diets (Colombo et al., 2014).

Several preventive zinc supplementation trials have also been conducted during pregnancy. The results have been inconsistent; only some have reported a positive effect of supplemental zinc on pregnancy outcome measures. In 2007, a meta-analysis showed a 14% reduction in infants born prematurely among zinc supplemented women, but no impact on infant birthweight, although there was a suggestion of a positive effect on birthweight in a subset of the women who were underweight or zinc-deficient (Hess and King, 2009). Two subsequent meta-analysis, one spanning five continents, of randomized controlled trials in which zinc supplements were given together with other micronutrient supplements, again indicated that only the effect on risk of preterm birth was significant, with no evidence of any effect on any parameter of fetal growth (low-birthweight, birthweight, length at birth, head circumference at birth) (Chaffee and King, 2012; Ota et al., 2015). More recently, a meta-analysis of studies of preventive zinc supplementation during pregnancy for three months or longer, reported no significant increase in birthweight and no decrease in the risk of low birthweight (Liu et al., 2018).

Zinc status is sometimes compromised in the elderly, probably because of low zinc intakes and age-related changes in physiological function (de Jong et al., 2001). As well, the presence of hypochlorhydria, malabsorption syndromes, and the use of certain medications may have exacerbating roles (Aggett and Comerford, 1995). There are several reports of improvements in some aspects of immune function in older adults, including nursing home elderly, following zinc supplementation (Bogden et al., 1994; Barnett et al., 2016). Other possible zinc-related degenerative changes in the elderly that warrant further study include hypogeusia, delayed wound healing, anorexia, deterioration of glucose tolerance, and depression (Marcellini et al., 2006; Levenson, 2006; Lin et al., 2017). For example, correlations between zinc status (based on serum zinc concentrations) and symptoms of depression have been observed in observational studies in elderly women (Marcellini et al., 2006).Improvements in patients with depression given supplemental zinc have also been reported, especially when the supplemental zinc is used in combination with antidepressant drug therapy (Ranjbar et al., 2014).

24c.1.5 Food sources and dietary intakes

There is no functional reserve or body store of zinc, as noted earlier, so an adequate regular supply of readily absorbable zinc is required. Hence, individuals consuming a usual diet low in zinc and/or with poor zinc bioavailability may be nutritionally zinc deficient. Food sources of readily absorbable zinc include organ meats, red meat, poultry, fish, and shellfish; oysters are one of the richest food sources of absorbable zinc. In general, the amount of zinc is higher in dark red meat than in white meat. Starchy roots, tubers, fruits, and vegetables have a low zinc content. Zinc is less readily available in whole grain cereals, nuts and legumes: their zinc content varies, depending on the zinc content of the soil or fertilizer treatment, growing location, and processing methods (e.g., milling of cereals). Loss of zinc from most foods during cooking is minimal.

Several dietary components have been investigated as potential modifiers of zinc bioavailability. A systematic review, however, concluded that for many of these potential modifiers, data were insufficient to accurately determine their impact on zinc absorption (Bel-Serrat et al., 2014). A notable exception is phytic acid (myoinositol hexakisphosphoric acid (InsP6) and its associated magnesium, potassium, and calcium salts — termed “phytates” — potent inhibitors of zinc absorption, as noted earlier. In general, phytic acid and its salts are found in high concentrations in unrefined cereals, legumes, nuts and oil seeds. Concentrations vary widely depending on the botanical variety, environmental or climatic growing conditions, the use of phosphate fertilizers, and the stage of maturation. The highest levels are reached at seed maturity. During some food-processing, preparation and cooking practices, such as milling / pounding, soaking, germination / malting, mixtamalization, or fermentation, phytate is dephosphorylated to the lower myo-inositol phosphate forms (IP1 to IP4) that don't inhibit zinc absorption (Gibson et al., 2018).

The inhibiting effect of phytate on zinc absorption follows a dose-dependent response. Hambidge and colleagues have developed a trivariant model to predict the inhibitory effect of different levels of phytate on zinc absorption as a function of zinc intake (Miller et al., 2007). Development of this model led the European Food Safety Authority (EFSA) to generate dietary zinc requirements for adults based on four levels of dietary phytate (EFSA, 2014). When daily phytate intakes in adults reach 1200mg/d, a level surpassed in some low income countries with staple diets of unfermented cereals and/or legumes, the average zinc requirement for adults, compiled by EFSA (2014) nearly doubles (Gibson et al., 2019). Values for the phytate content of raw and processed plant-based staples are available in the FAO / INFOODS / IZiNCG Global Food Composition Database for Phytate.

Molar ratios of phytate-to-zinc of individual foods or whole diets can be used to estimate the likely proportion of zinc absorbed. Diets with a phytate-to-zinc molar ratio of more than about 15 have poor zinc bioavailability, whereas those with ratios between 5 and 15 are said to have medium bioavailability, and those with ratios less than 5 have good bioavailability (WHO/FAO, 2004).

In the past, millimolar ratios of phytate × calcium : zinc were also used to predict zinc absorption because dietary calcium was said to influence the inhibitory effect of phytate on zinc bioavailability. The use of this ratio has now been discontinued in the absence of any demonstrated effect of calcium on zinc absorption from diets adequate in zinc, irrespective of whether intakes of dietary phytate were high or low (Hunt and Beiseigel, 2009). Whether calcium has an adverse effect in phytate-containing diets low in zinc is uncertain (King et al., 2015). Earlier, the amount and type of dietary protein was said to enhance zinc absorption but more recent data have not confirmed this (IZiNCG, 2004; Miller et al., 2013).

A competitive interaction between iron and zinc may occur, depending on the form of the iron, as well as the conditions and levels under which both the iron and zinc are given (Lönnerdal, 2000). A decrease in zinc absorption may occur when the ratio of iron to zinc is very high (i.e., > 25:1) and when the iron is administered in water and fed under fasting conditions. Hence, women taking high-dose prenatal iron supplements (60mg elemental iron/day) may require additional zinc (Caulfield et al., 1999). For example, supplementation with 60mg elemental iron in a multiple micronutrient supplement blunted the increase in serum zinc observed in Cambodian women who received the same supplement but with no iron (Holmes et al., 2019). However, this interaction is less apparent when the iron is provided in a complex food matrix (Olivares et al., 2012; Esamai et al., 2014).

In general, in omnivorous adult diets in affluent countries, meat and meat products supply the greatest amount of zinc, followed by cereals and dairy products (Briefel and Johnson, 2004; EFSA, 2014). However, interest in vegetarian diets has increased in these countries with high-phytate diets of cereals, legumes and nuts replacing meat and meat products (American Dietetic Association, 2003). Increasingly zinc fortified beverages, fortified cereals, and supplements are becoming important sources of dietary zinc, especially among young children in North America (Arsenault and Brown, 2003; Butte et al., 2010).

24c.1.6 Effects of high intakes of zinc

There are no reported cases of toxicity from excess intakes of dietary zinc, although toxicity arising from short-term exposure to very high levels of contaminant zinc (> 300ppm), from the improper storage of food or beverages in galvanized vessels, has been

reported (Brown et al., 1964). Signs and symptoms include dehydration, vomiting, electrolyte imbalance, abdominal pain, nausea, lethargy, dizziness, and muscular incoordination (Fosmire, 1990).

Adverse effects of excess zinc on copper metabolism have been described (Prasad et al., 1978a; Fischer et al.,1984; Samman and Roberts, 1988; Yadrick et al., 1989) with a decrease in the activity of erythrocyte Cu,Zn-SOD being the most consistent finding (Fischer et al.,1984).However, the clinical significance of a decrease in erythrocyte SOD activity is unknown.

High intakes of zinc have also been associated with reductions in serum HDL cholesterol levels in some studies (Hooper et al., 1980),. and detrimental effects on the immune system (Chandra, 1984). Hence, excessive self-supplementation with zinc may have an adverse effect on health, although doses of 25–35mg/day zinc in adults do not appear to pose a health hazard (Smith, 1994). In Europe the zinc content of the most common single nutrient supplement is 30mg per capsule (range 15–50mg), and 10–15mg (range 2–20mg) for multi-nutrient supplements(EFSA, 2002).

The Tolerable Upper Intake Level (UL) for zinc set by the U.S. Food and Nutrition Board for adults and pregnant and lactating women > 19y, is 40mg/d, a level based on the adverse effects of high doses of supplemental zinc (50mg Zn/d) on copper status biomarkers, notably reductions in the activity of erythrocyte Cu,Zn-SOD. For infants the corresponding UL is 4mg, with adjustments for children based on body weight (IOM, 2001).Concern has been raised that the UL set by U.S. Food and Nutrition Board for preschool children (i.e., 7mg/day) in the US and Canada may be too low as many have dietary zinc intakes that exceed the UL due to their consumption of zinc-fortified beverages, fortified ready-to-eat breakfast cereals, and zinc supplements (Arsenault and Brown, 2003; Butte et al., 2010). Sensitive biomarkers of copper status were reportedly unchanged by supplementation with 5, 10, or 15mg of zinc daily for four months in healthy boys in a double-blind, placebo-controlled randomized trial(Bertinato et al 2013). EFSA (2002) has set a lower level for the UL for adults and pregnant and lactating women (25mg Zn/d), although their UL for preschoolers 1–3y is the same as that of U.S. Food and Nutrition Board (i.e., 7mg/d).

24c.1.7 Biomarkers of zinc status

Zinc is classified as a type II nutrient (Golden, 1995) because it is essential for multiple metabolic functions, as discussed earlier. As a consequence, in zinc deficiency the physiological effects that arise are associated with a number of diverse biochemical changes, making it difficult to identify biomarkers of zinc nutrition (King, 2011). In contrast, reductions in tissue nutrient concentrations of type I nutrients, such as iron, iodine, and vitamin A, arising from deficits in dietary intakes, cause a decline in one or more specific functions which can be readily identified using selected biomarkers (King, 2011).

To identify populations at elevated risk of zinc deficiency, three indicators were recommended by WHO / UNICF / IAEA / IZiNCG: (a) prevalence of usual zinc intakes below the estimated average requirement, (b) percentage with low plasma/serum zinc concentrations, and (c) percentage of children aged < 5y who are stunted (i.e., HAZ < –2SD) ; (de Benoist et al., 2007), These recommendations were endorsed by the BOND Expert Zinc Panel (King et al., 2015). For details see IZiNCG Technical Brief No. 1 (2007).

For each indicator, a “trigger level” is given for the prevalence considered indicative of elevated risk and of public health concern, at which level an intervention to improve population zinc status is warranted. Reference limits for plasma/serum zinc (depending on age, sex, time of day, fasting status) are available to identify populations at risk of zinc deficiency, provided the correct protocols are followed for the collection, separation, and analysis of plasma zinc. Assessment of the functional indicator — impaired linear growth — should be made in combination with at least one of the other indicators in view of its poor specificity (King et al., 2015). In the future, a model comprising key metabolic indicators, serum albumin, and a biomarker of inflammation may be developed for population zinc assessment (King, 2018).

Assessment of zinc status at the individual level is particularly difficult because serum or plasma zinc concentrations are homeostatically controlled and do not decline with marginal zinc intakes. Hence, serum or plasma zinc is only useful to measure zinc status in individuals with either a very low or high supply of dietary zinc. Instead, the recommended approach for individuals combines a medical history, dietary and clinical assessment, with plasma or serum zinc providing only supplementary data (King et al., 2015).

The BOND Zinc Expert Panel classified the other possible zinc biomarkers into three groups: (1) potentially useful (2) emerging

biomarkers and (3) those deemed not useful. These are shown below in Box 24c.1.

Box 24c.1 Possible Zinc Biomarkers. From (King et al., 2015).

- **Potential Biomarkers** that show promise, but data are insufficient to establish specific cutoffs indicating zinc inadequacy in populations: Hair zinc, Urinary zinc, Neurobehavioral function
- **Emerging Biomarkers** for which there is some theoretical basis for a relation to zinc intake or status, but testing is insufficient to confirm the relation: Nail zinc, Zinc-dependent proteins (i.e. cellular metallothionein, metallothionein and zinc-transporter expression), Oxidative stress and DNA integrity, Zinc kinetics , Taste acuity
- **Not useful Biomarkers** that do not relate consistently to zinc intake or status: Zinc-dependent enzymes, Erythrocyte and leukocyte zinc

This page titled 28.1: Zinc (24c.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.2: Dietary Zinc Intake (24c.2)

Assessment of dietary zinc intakes is the best method for estimating zinc exposure in individuals and populations. To determine dietary zinc intakes, a quantitative dietary method such as a weighed or estimated food record or a validated 24-h recall must be used. Details of the procedures to estimate dietary zinc intakes using a modified 24-hr recall specifically for use in lower income countries are available in a Dietary Technical Monograph (#8) (Gibson and Ferguson, 2008).

Information on dietary zinc intakes can be used for several purposes. At the population level, data can be used to: estimate the prevalence of inadequate zinc intakes, classify subpopulations at elevated risk, design and monitor zinc intervention programs, and identify dietary patterns that contribute to inadequate zinc intakes. At the individual level, those at risk of inadequate intakes can be identified for dietary counseling. Tools have been developed to facilitate the collection of dietary zinc intakes and analyze the data to assess dietary adequacy (IZiNCG Technical Briefs) No.3 and No.7.

Inadequate intakes of dietary zinc can arise from low intakes of zinc *per se,* poor bioavailability, or a combination of these dietary factors, as noted earlier. In low income countries, where unrefined cereals and legumes containing high levels of phytate are often the major source of energy and zinc, inadequate intakes of bioavailable forms of zinc are the most likely cause of zinc deficiency. Other diets based on starchy roots or tubers have a low total zinc and phytate content. Values for the phytate content of raw and processed plant-based staples are available in the FAO / INFOODS / IZiNCG Global Food Composition Database (Infoods Phytate Database). When collecting dietary intake data for zinc, intakes of dietary phytate should also be determined, whenever possible.

Data on both zinc and phytate intakes permit the calculation of phytate-to-zinc molar ratios from which an estimate of the bioavailability of zinc can be obtained, as noted in Section 24c.1.5

24c.2.1 Measurement of zinc intake for populations

The recommended dietary indicator to identify a population or population subgroup at elevated risk of zinc deficiency is the prevalence of usual zinc intakes below the estimated average requirement (King et al., 2015). Because information on usual zinc intakes is required, food intakes must be measured on two non-consecutive days or three consecutive days. It is preferable that multiple days of intake data are collected for all individuals in the population. However, if this is not possible, then at least two non-consecutive days of dietary intake data from a sub-sample(40–50) of individuals per stratum should be collected. Figure 24c.4

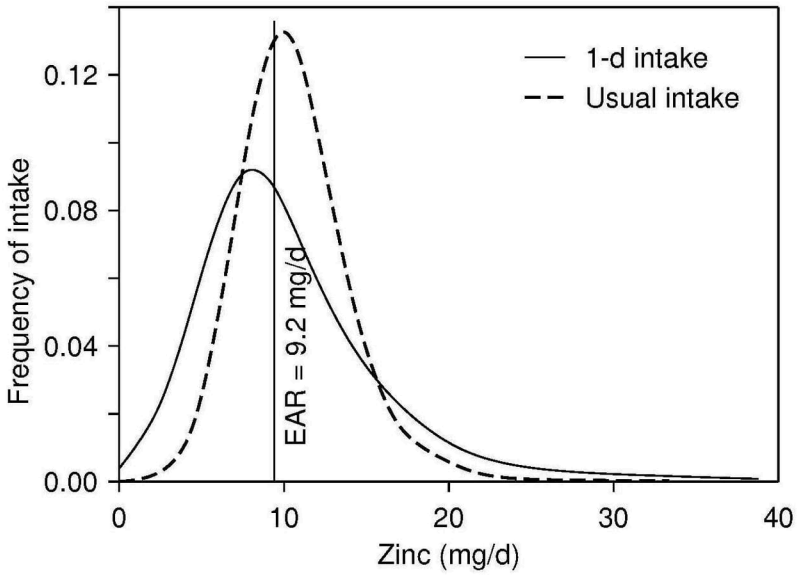


Figure 24c.4 Estimates of usual intake distribution for zinc for New Zealand adults obtained from 24h recall data and adjusted with replicate intake data using the refined NRC method. The y-axis (frequency of intake) shows the likelihood of each level of intake in the population. EAR, Estimated Average Requirement. Redrawn from Gibson et al. (2003).

These data permit an estimate of the day-to-day variation in zinc intake within one individual (i.e., within-person variation) to be estimated, allowing a distribution of usual zinc intakes for the population group to be generated using specialized software. Finally, the Estimated Average Requirements (EARs) for zinc, specific for age, sex, physiological status, and in some cases the phytate-to-zinc molar ratio of the diet, are needed to calculate the prevalence of inadequate intakes. If quantitative data on phytate intakes are not available, then the phytate-to-zinc molar ratio can be estimated based on the diet type. Figure 24c.4 shows an example comparing the adjusted distributions of usual zinc intakes with the observed one-day zinc intakes for New Zealand adult females aged 19–50y, using the program developed by Iowa State (ISU) and implements the ISU method (Nusser et al., 1996) for the adjustment. The adjustment process used yields a distribution with reduced variability that preserves the shape of the original observed distribution (Gibson et al., 2003). The example also shows that in this particular case, adjusting the distribution significantly reduces the proportion of individuals considered to have intakes below the EAR. An elevated risk of zinc deficiency in the population is said to exist when 25% or more of the population have zinc intakes less than the EAR.

The five main steps required to calculate this recommended dietary indicator are summarized in Box 24c.2.

Box 24c.2 Steps to determine the dietary intake at the population level

- Select a representative sample of the population
- Measure food intake by using either a 1-day weighed or estimated food record or a validated 24-h recall, preferably on at least two non-consecutive days for each individual or for at least a subsample of individuals in the population (40-50 individuals for each life-stage group with a different zinc requirement).
- Calculate zinc and phytate intakes and dietary phytate-to-zinc molar ratios for each individual by using an appropriate food composition database for the country. For non-pregnant and non-lactating adults, the amount of zinc likely to be absorbed may also be calculated.
- Adjust the distribution of observed zinc intakes to represent usual zinc intakes by removing the variability introduced by day-to-day variation in an individual's zinc intake.
- Select the appropriate EAR and apply the EAR cutoff method to estimate the prevalence of usual zinc intakes below the EAR. Alternatively, in countries where local food composition values for zinc and phytate are not available, collect 24-h duplicate diet composites from each individual for the chemical analysis of zinc and phytate, again with repeats on ≥40–50 individuals per stratum as described above (IZiNCG Technical Briefs) No.3.

24c.2.2 Interpretation of zinc intakes for populations

Several expert groups have set dietary zinc recommendations that include EARs for zinc, although discrepancies exist, depending on the sources of the data, the concepts and methods used, as described in Gibson et al. (2016). For example, the EARs for zinc set by the United States and Canada are based on a fixed adjustment for zinc absorption from habitual diets for both children and adults (Food and Nutrition Board, 2001). In contrast, the European Food Safety Authority (EFSA, 2014) have generated dietary zinc requirements based on four levels of dietary phytate intake for adults (> 18y) and said to cover the average phytate intakes in European populations: (300mg/d; 600mg/d; 900mg/d; 1200mg/d). In view of the uncertainty about whether recommendations for dietary zinc based on phytate intakes can also be made for young children, the EFSA EARs for infants and children are not adjusted for phytate.

The prevalence of usual zinc intakes below the EAR is simply estimated by calculating the percentage of individuals within a specific life-stage group with usual intakes below the respective EAR. An elevated risk of zinc deficiency in the population is said to exist when 25% or more of the population have zinc intakes less than the EAR. However, to assist with the interpretation of the risk of zinc deficiency, it is recommended that estimates of the adequacy of zinc intakes be combined with biochemical data on serum zinc concentrations (IZiNCG Technical Brief) No.3.

In cases where the amount of zinc likely to be absorbed is calculated for non-pregnant and non-lactating adults ≥ 19y using the updated trivariate saturation response model of Miller et al. (2007), then the prevalence of absorbable zinc intakes below the appropriate physiological requirement for absorbed zinc should be calculated. For IZiNCG, these requirements are 1.86mg/d for adult women and 2.69mg/d for adult men ≥ 19y of age. For further details of this calculation see (IZiNCG Technical Brief) No.3.

24c.2.3 Measurement and interpretation of usual zinc intakes for individuals

Usual intakes of dietary zinc at the individual level can also be determined using quantitative dietary methods such as a weighed or estimated food records or validated 24-h recalls, provided a large number of measurement days per individual are conducted. Details on how to define the desired number of replicate days per individual are available in Gibson and Ferguson (2008). Alternatively, a dietary history or validated semi-quantitative food frequency questionnaire can be used to provide retrospective information about usual food consumption patterns, preferably over more than one month; details are available in Thompson et al. (2015). Assessment of the intake of zinc-rich foods (e.g., meat, poultry, fish), zinc-fortified foods, and high-phytate foods (e.g., unrefined cereals, nuts, and legumes) can be used to predict the intake of absorbable zinc (King et al., 2015). Inferences can be made about the adequacy of the zinc intake of an individual by comparing the difference between the estimate of the usual zinc intake of the individual and the respective values for the EAR and Recommended Nutrient Intake (RNI) for zinc. If the usual intake of the individual is greater than the RNI, then there is a high level of confidence that the usual intake is adequate, whereas if the usual intake is between the EAR and the RNI, the usual zinc intake probably needs to be improved, and if the usual intake is less than the EAR, then the usual intake very likely needs to be improved. Details on the guidelines for the qualitative interpretation of individual intakes are available from the Institute of Medicine (2000).

This page titled 28.2: Dietary Zinc Intake (24c.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.3: Serum zinc (24c.3)

Serum and plasma zinc concentrations can be used interchangeably because differences between these two sources of circulating zinc are very small (English and Hambidge, 1988). Fasting serum zinc concentrations are homeostatically controlled within fairly narrow limits (about80–100µg/dL (12–15µmol/L). Only 0.1% of the total body zinc is present in the serum, whereas 3% and 63% are in the liver and skeletal muscle, respectively (Table 24c.1). Of the zinc in whole blood, from 12% to 22% is in the serum; the remainder is within the erythrocytes. Zinc is transported in the serum bound principally to albumin (70%), so conditions that alter serum albumin levels will, in turn, affect serum zinc concentrations. The remaining zinc in serum is tightly bound to α_2-macroglobulin (18%), and the rest is bound to other proteins such as transferrin and ceruloplasmin and to amino acids, primarily histidine and cysteine (Cousins, 1985).

Serum zinc concentration is considered the best available biomarker of zinc exposure, status, and risk of zinc deficiency at the population level(de Benoist et al., 2007; Lowe et al., 2009; Hess et al., 2007). Evidence for this recommendation is derived from investigations on the following: (1) the effect of dietary zinc restrictions and repletion on serum zinc concentrations; (2) the effects of zinc supplementation on serum zinc; (3) the association between serum zinc and clinical signs of zinc deficiency; and (4) a comparison of initial serum zinc concentrations between individuals who do or do not show a functional response to the correction of zinc deficiency (King et al., 2015).

In brief, in severe dietary zinc restriction (i.e., < 1mg/d) in previously healthy adults, there is a sharp decrease in serum zinc concentrations, with values returning to baseline levels within 1–2 weeks of zinc supplementation Figure 24c.5.

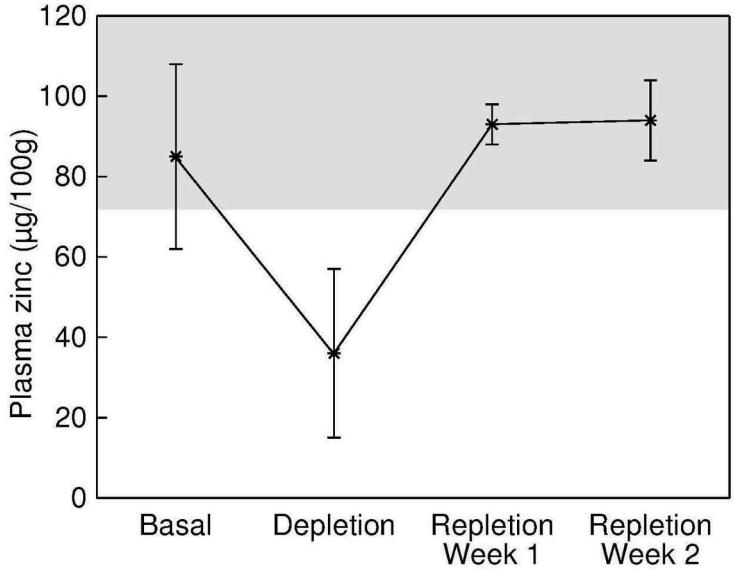


Figure 24c.5. Changes in plasma zinc (mean±SD) at the end of basal, depletion, and repletion periods. The shaded area represents the normal range. From Baer and King(1984).

In more moderate dietary zinc restrictions (3–5mg/d),however, reductions in serum zinc only occur after prolonged periods of a restricted diet or when accompanied by high phytate intakes (Gibson et al., 2008).

Supplemental zinc consistently and rapidly increases serum zinc in both children and adults regardless of initial serum zinc concentrations (Hess et al., 2007; Wessells et al., 2020). Of interest is the finding that serum zinc responds less to additional zinc provided in food than to additional zinc supplied as a supplement administered between meals. It is possible that absorbed zinc may be metabolized differently when consumed in food or as a supplement (King et al., 2015).

Low serum/plasma zinc has been related to clinical signs of zinc deficiency based on data from experimental zinc-depletion / repletion studies and case reports of acrodermatitis enteropathica. From these data, the sensitivity and specificity of a plasma zinc below 50µg/dL (7.65µmol/L) for detecting clinical signs of deficiency was 82% and 92%, respectively (Wessells et al., 2014).

Finally, if the initial serum zinc is very low, a functional response to zinc supplementation may occur. Functional responses may include increases in growth (including linear growth, weight gain, and accrual of fat-free mass) and decreased morbidity from diarrhea and respiratory infection. Of these responses, IZiNCG has selected linear growth outcome as the preferred functional response based on height- or length-for-age (de Benoist et al., 2007).

Serum or plasma zinc has been used to assess the risk of zinc deficiency in several national surveys. The population is considered to have an elevated risk of zinc deficiency of public health importance when a particular population has more than 20% of individuals with a serum zinc concentration below the age and sex-specific cutoffs. In a review of national surveys in twenty low- or middle-income countries the prevalence of low plasma zinc was more than 20% in 68% of the children and 93% of the women (Hess, 2017), compared to the US NHANES 2011–2014 survey where only 3.8% of children (6– < 10y) and 8.2% of females (≥ 10y) had low serum zinc concentrations (Hennigar et al., 2018).

In general, national surveys (Gibson et al., 2011; Engle-Stone et al., 2014; Hennigar et al., 2018) have not shown a positive relationship between individual dietary zinc intakes and serum zinc concentrations. Meta-analyses conducted on adults (Lowe et al., 2012), pregnant and lactating women (Moran et al., 2012a), and children (Moran et al., 2012b), showed only small changes in serum zinc when dietary intakes were doubled. These findings are not unexpected in view of the strong homeostatic mechanisms that prevent changes in serum zinc when zinc levels in the diet fluctuate. King (2018) has emphasized that doubling dietary zinc intake in adults only results in a 6% change in serum zinc concentrations, a difference within the margin of error in measuring serum zinc. Among some population groups in whom high-phytate cereals and legumes replace animal source foods as the major source of dietary zinc (e.g., vegetarians), significant inverse relationships between serum zinc and dietary phytate : zinc molar ratios have been observed (Donovan and Gibson, 1995; Cantoral et al., 2015).

For an individual, serum zinc concentrations are not a reliable indicator of zinc status due to both the strong homeostatic control mechanisms and the many factors other than zinc intake that can independently influence serum zinc concentrations (see below). This means that serum zinc concentrations may be within the normal range despite the presence of mild zinc deficiency generating growth retardation, loss of appetite, or impaired immune function. Currently for individuals the recommended approach combines a thorough medical history of risk factors for zinc deficiency, a qualitative dietary assessment to identify an inadequate dietary intake, and a clinical assessment, with plasma or serum zinc providing only supplementary data (King et al., 2015).

In individuals with more severe zinc deficiency, with manifestations such as skin lesions, hair loss, diarrhea, delayed sexual maturation, impotence, hypogonadism in males, and eye lesions, serum zinc concentrations are usually low. For example, markedly low serum zinc concentrations are present in individuals with the genetic disorder acrodematitis enteropathica (Wessells et al., 2014), in patients receiving TPN unsupplemented with zinc (Arakawa et al., 1976), and in those individuals who have developed clinical signs during experimentally-induced zinc deficiency (King et al., 2015).

24c.3.1 Factors affecting serum zinc

Age-related changes in serum zinc concentrations (Table 24c.3) may be apparent, with levels increasing with age in some Pilch and Senti, 1984. Villalpando et al., 2003; Gibson et al., 2011), but not all (Hennigar et al., 2018) national surveys and in some smaller studies (Rükgauer et al., 1997). In the 2011–2014 US NHANES III survey, where serum zinc was determined in males and females aged ≥ 6y, only the serum zinc concentrations for the males (not the females) increased from the ages of6–30y, declining thereafter

Table 24c.3 Median ±SE serum zinc levels of males and females in the United States by age. Data from the US NHANES 2011–2014 (For conversion to µmol/L, divide by 6.536)

Age (years)	Male serum zinc (µg/dL)	Female serum zinc (µg/dL)
6–8	80.8±1.4	77.7±0.9
9–13	81.7±1.2	81.1±1.6
14–18	85.5±2.3	81.3±1.9
19–30	84.9±1.4	77.9±1.3
31–50	84.4±0.7	79.6±0.7

Age (years)	Male serum zinc (µg/dL)	Female serum zinc (µg/dL)
51–70	82.4±1.9	79.8±1.5
≥71	78.7±2.3	80.5±1.5

Gender may influence serum zinc concentrations. During infancy and early childhood, boys tend to have lower serum zinc levels than girls (Smit-Vanderkooy and Gibson, 1987; Cavan et al., 1993a; Gibson et al., 2011). In the US NHANES III2011–2014 survey males from 6y to adults ≥ 71y had higher serum zinc concentrations than females (Table 24c.3), a relationship that may be related to differences in body size and lean body mass.

Diurnal variation in serum zinc concentrations has been reported. Levels are higher in serum in the morning, regardless of fasting status, compared to the afternoon Pilch and Senti, 1984. In the US NHANES III 2011–2014 survey, serum zinc concentrations were 9% lower in the blood samples drawn in the afternoon (Hennigar et al., 2018),(Table 24c.4)

Fasting status also markedly affects serum zinc concentrations. During a usual overnight fast of ≥ 8h, serum zinc levels increase, generally reaching a peak just before breakfast. Fluctuations in serum zinc also occur throughout the day, mainly in response to the consumption of meals. Serum zinc levels decline progressively for several hours after each meal, before rising prior to the next meal (King et al., 1994). Hence, in addition to the time of day, fasting status and meal status should also be carefully controlled during the collection of blood samples for serum / plasma zinc to avoid confounding the results. Alternatively, the time of day and the time elapsed since the previous meal can be recorded and taken into account statistically during analysis (Diana et al., 2017) (Table 24c.4).

Table 24c.4 Median serum zinc for different collection times from NHANES III. Data from the US NHANES 2011–2014 (Hennigar et (For conversion to µmol/L, divide by 6.536)

All participants Time/fasting status	n	Median serum zinc (µg/dL ± SE)
Morning, fasting	1423	86.8 ± 0.8
Morning, nonfasting	89	85.0 ± 3.1
Afternoon	1124	76.9 ± 0.9
Evening	489	73.8 ± 0.9

Hemolysis either *in vivo* or *in vitro* has the potential to increase serum zinc concentrations because the concentration of zinc in erythrocytes is about 10–20 times higher than in serum or plasma. Hemolysis may be particularly important in cases of zinc deficiency, when red cell fragility increases (Bettger et al., 1978). Killilea et al. (2017) suggest that a 5% increase in plasma or serum zinc concentration occurs with each 1g hemoglobin/L in plasma or serum, a concentration that can be identified by chemical hemoglobin assays or by matching to a color scale. A threshold of 1g hemoglobin/L is recommended for measurements of plasma or serum zinc to avoid significant increases in zinc caused by hemolysis.

Storage of blood samples prior to separation affects serum zinc concentrations. Long intervals prior to separation are associated with progressively increasing serum zinc concentrations as zinc is released from platelets(English and Hambidge, 1988). To avoid this increase, serum or plasma should be separated rapidly from the red blood cells within 30–40 minutesafter collection. If this is not feasible, the blood should be held at 2–10°C for no more than 24 hours to limit movement of cellular zinc into the plasma or serum (IZiNCG Technical Brief No.2, 2007).

Prolonged use of a tourniquet may result in an increase in serum or plasma zinc levels. This occurs as a result of increased intravascular pressure caused by venous occlusion, which may cause movement of fluid into the interstitial space, thus increasing the zinc concentration(Juswigg et al., 1982). Hence, tourniquet occlusion should only be used for about one minute (King et al., 2015).

Acute and chronic infection and systemic inflammation leads to serum or plasma zinc values that are spuriously low because of the transfer of zinc from the blood to the liver. This redistribution of zinc is caused by hepatic synthesis of metallothionein activated by the release of cytokines during the acute phase response (Raiten et al., 2015). Both the stage and severity of the infection influence the change in plasma zinc, but these changes do not correspond to a change in zinc nutrition. To adjust for the presence of systemic inflammation, two inflammatory biomarkers (C-reactive protein (CRP) and α-1-acid glycoprotein (AGP)) should be measured along with serum or plasma zinc, allowing a regression-based correction to be applied (McDonald et al., 2020).

Evaluation of zinc intervention programs requires scheduling the end-line blood collection before the end of the intervention because serum zinc concentrations rapidly return to baseline levels after the withdrawal of supplementation (King et al., 2015).

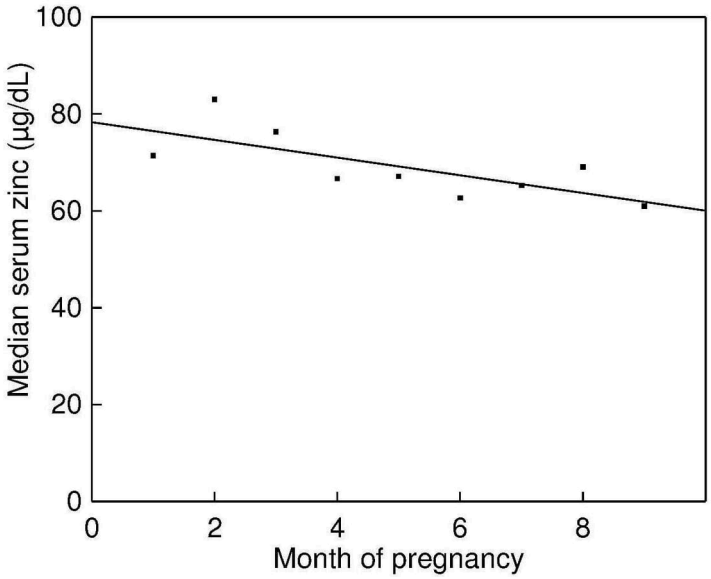


Figure 24c.6. Median serum zinc concentrations during pregnancy at 1-mo intervals with the line of linear regression. From Brown et al., (2004)

Pregnancy is associated with decreases in serum zinc concentrations irrespective of zinc intake Donangelo and King, 2012). The decline is evident at two months of gestation and at this early stage is attributed to hormonal changes, whereas the later decline is linked to plasma volume expansion (King, 2018). At the end of pregnancy, serum zinc concentrations are about 20–25% lower than pre-pregnancy values.(Hennigar et al., 2018). In the US NHANES III 2011–2014 survey, the geometric mean ± SD serum zinc for pregnant women (n=34) was69.1±12.0µg/dL vs. 79.1±1.9µg/dL (10.6±0.3 vs. 12.1±0.3µmol/L) for the age-matched non-pregnant women population; data by month of pregnancy are not available. The median serum zinc concentrations by month of pregnancy, drawn from the NHANES II data is shown in Figure 24c.6.

Hypoalbuminemia (serum albumin ≤ 3.5g/dL) is associated with an increased risk of low serum zinc concentrations because about60–70% of the serum zinc is bound to albumin (Cousins, 1989). conditions such as inflammation and infection, and disease states such as alcoholic cirrhosis and protein-energy malnutrition, result in a decline in serum zinc concentrations concurrent with hypoalbuminemia which are probably independent of an individual's zinc status (King, 2018).Not surprisingly, in the US NHANES III 2011–2014 survey, low serum albumen concentrations were associated with an increased risk of low serum zinc (Hennigar et al., 2018).

Estrogen-containing preparations, such as oral contraceptive agents and other hormones, when used, may lead to markedly lower serum zinc oncentrations (Swanson and King, 1982), depending on the hormone dosage (Kamp et al., 2011). (Table 24c.5).

Table 24c.5 The effect of oral contraceptive (OC) use on serum zinc and alkaline phosphatase concentrations. [a]After correction for a

Participants (n)	Mean serum zinc (µmol/L)	Mean serum alkaline phosphatase
Non-users (202)	12.11	22.8
OC Users (128)	11.81	19.5
Probability of significant diff.[a]	0.05	0.01

In the US NHANES III survey, however, there were no differences in serum zinc concentrations between oral contraceptive users and nonusers, perhaps because of the now decreased hormone dosage (Hennigar et al., 2018). The IZiNCG Technical Brief No.2 (2007) advises the collection of data on the use of oral contraceptive agents and other hormones in studies measuring serum zinc concentrations.

Rapid synthesis of tissues during growth may lead to a fall in serum or plasma zinc. This effect has been reported in children during the anabolic phases of recovery from malnutrition (Golden and Golden, 1981) and during the rapid growth that occurs in preterm infants (Altigani et al., 1989). The decline in plasma zinc probably arises from the increased uptake of zinc from the exchangeable zinc pool in the plasma, induced by the increased demands for zinc for growth (Aggett and Comerford, 1995).

Weight loss or starvation releases zinc from muscle breakdown, resulting in transient increases in serum zinc concentrations (Henry and Elmes, 1975; IZiNCG Technical Brief No.2, 2007).

Anemia has been associated with low serum zinc concentrations in a few studies in children in both high-resource (Cole et al., 2010; Houghton et al., 2016) and low resource (Wieringa et al., 2016) settings, in pregnant women in rural Ethiopia (Gibson et al., 2008), and in females in the2011–2014 US NHANES III survey (Hennigar et al., 2018). Several mechanisms may explain the association of serum zinc and hemoglobin concentrations (King, 2018): zinc-dependent enzymes are needed for heme synthesis, and zinc may also stabilize the erythrocyte membrane, thereby protecting the membrane from degradation during oxidative stress (O'Dell, 2000).

Malabsorption syndromes and inflammatory bowel diseases result in low serum or plasma zinc concentrations arising from alterations in the integrity of the mucosal cells and a reduction in zinc absorption (King and Cousins, 2014).

Environmental enteric dysfunction is known to impair zinc absorption and increase intestinal endogenous zinc losses (Lindenmayer et al., 2014) that together result in low serum or plasma zinc concentrations (Manary et al., 2010).

Sickle cell disease has been associated with low plasma zinc concentrations in children and young adults (Abshire et al., 1988). This has been attributed to increased urinary zinc excretion, possibly mediated by increased zinc mobilization from bones due to recurrent bone ischemia (Schimmel et al., 2016).

High levels of supplemental iron beyond those usually consumed in the diet may reduce serum zinc concentrations due to a decrease in zinc absorption. The mechanism is uncertain. Iron and zinc are thought to compete for intestinal absorption via the shared divalent metal transporter 1 (DMT1) and/or another common pathway in the apical membrane of the intestinal cell (Gunshin et al., 1997). The adverse effect on zinc absorption, however, is less when lower amounts (≤ 10mg) of iron supplements are given (Fischer Walker et al., 2005) or when both the iron and zinc are provided in a complex food matrix (Olivares et al., 2012; Esamai et al., 2014).

24c.3.2 Interpretive criteria

Table 24c.4 (shown above) presents the medium serum zinc by fasting status and time of sampling for all NHANES III participants.

The International Zinc Consultative Group (IZiNCG) reanalyzed serum zinc data from NHANES II. In this reanalysis, data for participants with conditions significantly affecting serum zinc concentrations were excluded. These were individuals: with low serum albumin (< 3.5g/dL); with an elevated white blood cell count (> 11.500 cells per µL); using oral contraceptive agents, hormones or steroids; or experiencing diarrhea. The IZiNCG also took age, gender, fasting status (i.e., > 8h since the last meal), and time of day of the sample collection into account in the reanalysis; details are given in (Hotz et al., 2003) and (IZiNCG, 2004).

From these reference data on healthy individuals, IZiNCG defined statistically derived reference limits based on the 2.5th percentile of serum zinc concentration for males and females aged < 10y and ≥ 10y (by fasting status) and time of sampling. These are given in (Table 24c.6) for children < 3y. IZiNCG suggests that the cutoffs for children < 10y also be applied to children < 3y, until appropriate reference data are available (IZiNCG, 2004).

Table 24c.6 Suggested lower reference limits (2.5th percentiles) (µmol/L) for serum zinc concentration based on age group, sex, fas divide by 0.1530. Adult reference limits based on data from subjects 20y and older only. NA, not available. Data from Brown et al. (2

Collection time	Children < 10y	Males ≥ 10y
AM fasting	NA	11.3
AM other	9.9	10.7
PM	8.7	9.3

Only a small number of pregnant women were examined during NHANES II (n=61). The 2.5th percentile values for the first and

second plus third trimesters of pregnancy were calculated by IZiNCG and are 8.6 and 7.6µmol/L, respectively. No estimate for the 2.5th percentile serum zinc value for lactating women can be derived from the NHANES II survey results because of the limited sample size. Instead, at present the cutoff derived for nonpregnant women should be used for lactating women (IZiNCG, 2004). Further reference data are required to confirm and develop lower cutoffs for pregnant and lactating women and infants and children < 3y.

In addition to recording age group and sex, data on both fasting status (including the time interval since the last meal) and the time of day of the blood collection, should also be collected. The IZiNCG Technical Brief No.2 (2007) also recommends measuring an acute-phase protein indicative of infection or inflammation to assist in the interpretation of the serum zinc results. Ideally the two acute phase proteins recommended by WHO (2014) — C-reactive protein (CRP) and α1-acid glycoprotein (AGP) — should be measured as elevated levels could indicate underlying inflammation, which reduces serum/plasma zinc concentrations (Raiten et al., 2015). Hence, before applying the appropriate IZiNCG reference limits, serum or plasma zinc concentrations may need to be adjusted statistically for the time of day, the time interval since the previous meal (Arsenault et al., 2011), and inflammation (Macdonald et al., 2020), in order to provide a true estimate of the prevalence of low serum zinc concentrations. To ascertain whether adjustment for inflammation is warranted, correlation and decile analysis between plasma / serum zinc concentrations and CRP or AGP should be conducted. If the correlation is negative and significant, a visual inspection of the decile analysis should be undertaken to confirm whether there is an increasing prevalence of zinc deficiency by CRP or AGP decile (McDonald et al., 2020). If either of the latter is observed, then the regression correction approach recommended by the Biomarkers Reflecting Inflammation and Nutritional Determinants of Anemia (BRINDA) Project should be used to adjust the serum / plasma zinc concentrations affected by inflammation. In this approach the inflammatory biomarkers (CRP and AGP) are treated as continuous variables so that greater corrections can be applied when the inflammatory biomarkers indicate severe inflammation (Namaste et al., 2017; McDonald et al., 2020).

IZiNCG (Technical Brief No.2, 2007) has recommended that if more than 20% of the population (or population sub-group) has a serum zinc concentration (after appropriate adjustment) below the relevant reference limit (based on age, sex, time of day of the blood collection, and fasting status), the whole population (or sub-group) should be considered to have an elevated risk of deficiency of public health importance. This “trigger level”, however, can be modified depending on the available resources within a country to control zinc deficiency (King et al., 2015).

Not surprisingly, the cutoff associated with overt clinical signs of zinc deficiency (i.e. 50µg/L, 7.6µmol/L) (Wessells et al., 2014) is higher than the statistically derived IZiNCG reference limits presented in Table 24c.6 that are based on reference data from healthy individuals. As such, these IZiNCG reference limits provide a greater margin of safety for assessing an increased risk of zinc deficiency at the population level before the appearance of clinical signs (King et al., 2015).

24c.3.3 Measurement of serum zinc

Blood samples for serum or plasma zinc should be taken under carefully controlled, standardized conditions. Contamination from various sources such as preservatives, evacuated tubes, lubricants, anticoagulants, water, and rubber stoppers must be avoided. The reader is advised to consult the IZiNCG website for practical tips on the collection of blood samples for the measurement of serum or plasma zinc (IZiNCG Technical Brief No.2, 2007). For venipuncture blood samples, trace-element-free evacuated tubes with siliconized rather than rubber stoppers must always be used. Stainless steel or siliconized needles and Teflon or polypropylene catheters can all be used. For capillary blood samples, the use of polyethylene serum separators with polyethylene stoppers and olefin-oligomer is recommended (Iyengar, 1998). As noted earlier, fasting status, time of day of blood collection, time elapsed since the previous meal and since the blood sample was centrifuged, and presence of inflammation should all be recorded so that serum zinc values can be adjusted statistically as needed (Arsenault et al., 2011) prior to selecting the appropriate IZiNCG reference limits.

Box 24c.3. shows pre-collection considerations recommended by IZiNCG (2007) to minimize contamination during the collection of samples for serum / plasma zinc:

Box 24c.3 Pre-collection considerations to minimize contamination

- Obtain supplies of diposable polyethylene gloves, free of talc or other coatings, for those handling blood samples.
- Arrange for samples be processed in a laminar flow class 100 clean room, laminar flow hood, or otherwise in a clean, dust- and smoke-free laboratory.

- Check that stainless steel needles are available.
- Pre-screen anticoagulants, if required, for adventitious zinc contamination prior to use; zinc-free heparin is preferred.
- Pre-screen trace-element-free polyethylene evacuated tubes, stoppers, and serum separators for zinc contamination prior to use.
- Pre-screen polyethylene storage vials and transfer pipets for zinc contamination before use.
- Decontaminate all equipment to be used, with the exception of the prescreened disposable items. This can be achieved by careful washing (soak for 24h in ultrapure 10%–20% HCl or HNO_3 solution and then rinsing three or four times with distilled, deionized water)
- Cover or seal all materials and equipment to be used, prior to storage to avoid dust.

Once the blood has been collected, blood samples should be placed in a refrigerator or on ice and allowed to clot for 30–40 minutes, and then centrifuged to separate the serum or plasma. This approach minimizes any transfer of zinc from the cells to the serum or plasma. Alternatively, if centrifugation on site is not feasible, then the blood should be placed in a cool box or refrigerator immediately after collection where it can be held at 2–10°C for up to 24 hours prior to being centrifuged and the serum or plasma separated. After separation, the serum or plasma should be transferred to a trace-element free screw-top vial for storage either in a refrigerator (for up to several days) or frozen until analyses. Any obviously111 hemolyzed samples (i.e., > 1g hemoglobin/L) should be discarded (Killilea et al., 2017; IZiNCG Technical Brief No.xx, 2020).

Some investigators prefer to use plasma because it is readily separated, less susceptible to platelet contamination, and not subject to contamination from a reaming instrument (Smith et al., 1985). For plasma samples, the recommended anticoagulant is zinc-free heparin. Before using any other anticoagulant it should be screened for contaminant zinc. For more details, see King et al. (2015).

The most frequently used method for the analysis of serum zinc is flame atomic absorption spectrometry (AAS), using a direct method in which the samples are diluted by 5 to 10-fold in solvents such as 6% aqueous butanol or 10% aqueous propanol (King et al., 2015). Dilution of the serum reduces the viscosity of the sample, which minimizes both the matrix effect on the rate of aspiration and the tendency for the flame atomic absorption burner-head to become blocked. This may be especially a problem with plasma because of precipitates that form in these samples. Sometimes the serum or plasma samples are ashed at low temperatures prior to analysis by flame AAS. For very small samples, flameless AAS can be used. Increasingly, inductively coupled plasma (ICP) atomic emission spectrometry or ICP mass spectrometry are being used. For all methods, in-house quality controls and standard reference materials (SRMs) such as bovine serum (SRM 1598) from the National Institute of Standards and Technology should be used. A coefficient of variation of < 5% is attainable for zinc using flame AAS. Use of a trichloroacetic acid deproteinization technique is not recommended.

Box 24c.4 shows the optimal procedures for the collection and analysis of serum or plasma samples to avoid adventitious trace element contamination. as recommended by IZiNCG (2004).

Box 24c.4 Optimal collection and analysis procedures

- Arrange for the participant to be seated.
- Clean the participant's skin with alcohol at the site of the antecubital vein.
- Restrict occlusion of the participants's arm with tourniquet to about 1 min.
- Draw blood, using a stainless steel needle, and collect into trace-element-free evacuated blood collection tubes *without* anticoagulant for processing serum.
- Place blood sample immediately in a refrigerator or on ice and allow to clot for 30–40min.
- Centrifuge blood sample at 2000–3000g for 10–15 min.
- Separate the serum using a zinc-free tranfer pipet.
- Discard any obviously hemolyzed serum samples.
- Consider retaining frozen red-cell residue for other determinations.
- Store serum samples frozen unless they are to be analyzed immediately.
- For analysis, dilute the sample quantitatively 5 or10-fold in solvents such as 6% aqueous butanol or 10% aqueous propanol.
- As soon as possible, determine sample zinc concentration using AAS or an alternative with appropriate standard dilutions, along with in-house quality controls and standard reference materials.

Practical tips for implementing these procedures are provided in IZiNCG Technical Brief No.2 (2012). Serum or plasma samples for zinc analyses can be refrigerated (4°C) for 2–3 weeks prior to analysis. For longer storage periods, samples should be frozen at −25°C or below. To prevent dehydration during long-term storage, especially if frost-free freezers are used, serum or plasma samples should be stored together with ice cubes in sealed plastic bags.

This page titled 28.3: Serum zinc (24c.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.4: Linear growth (24c.4)

The limited sensitivity and specificity of linear growth and other similar “bioindicators” means linear growth must be measured alongside other biomarkers of zinc (i.e., plasma or serum zinc) and other growth-limiting nutrients to establish the role of zinc in poor growth. There is no pharmacological effect of zinc on growth in zinc-replete individuals and linear growth is considered the best functional bioindicator associated with the risk of zinc deficiency in populations (King et al., 2015). Other functional bioindicators known to be responsive to zinc supplementation (e.g., diarrheal episodes) are difficult to define in a standardized manner. Justification for the selection of linear growth as the zinc functional bioindicator is summarized below in Box 24c.5. (Fischer -Walker and Black, 2005).

Box 24c.5 Justification for linear growth as the zinc functional bioindicator

- Low height- or length-for-age is often responsive to zinc supplements.
- Standardized methods exist to measure height or weight that are widely used.
- Reference data (i.e.,WHO growth reference data) (WHO , 2006) are available for interpretation.
- Linear growth is likely to be the primary response to an increased intake of absorbable zinc whereas weight gain is likely to occur as a result of increased linear growth.

In a study by Wessells and Brown (2012), the prevalence of inadequate zinc intakes in 138 low- and middle-income countries was estimated from food balance sheet data. In these countries the prevalence of stunting in children less than 5y was positively correlated with the estimated prevalence of inadequate zinc intake (r=0.48. P=0.0001) as shown in Figure 24c.7.

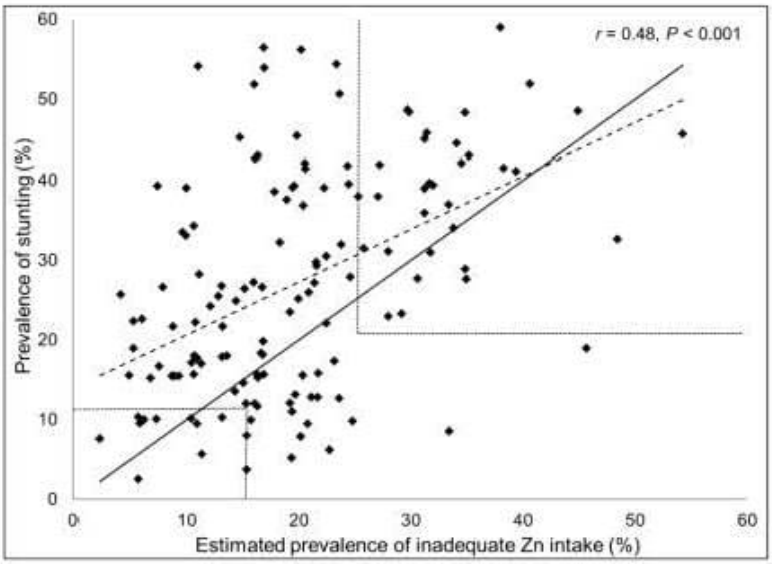


Figure 24c.7 Relationship between the estimated prevalence of inadequate zinc intake and the prevalence of childhood stunting. Stunting data (low height-for-age) are for children less than 5y in138 low- and middle-income countries. The solid line represents the line of identity (intercept=0, slope=1). The dashed line represents the best-fit regression line. Dotted lines demarcate countries with a high risk of inadequate zinc intake and where the prevalence of stunting is > 20%. From Wessells & Brown (2012).

The figure also defines an at-risk group of 32 countries where the estimated prevalence of inadequate zinc intake is > 25%, and the prevalence of stunting is > 20%.Nevertheless, both inadequate intakes of zinc and stunting only provide suggestive evidence of zinc deficiency, and the multi-factorial causes of childhood stunting may be responsible, at least in part, for the marked variability around the “best-fit” regression (dashed) line.

24c.4.1 Interpretive criteria

The percentage of children < 5y with height- or length-for-age less than –2SDs below the age-specific median of the WHO Multicentre Growth Reference Study (MGRS) population is recommended for assessing the zinc status of populations (WHO, 2006). Risk of zinc deficiency is considered to be of public health concern when the prevalence of low height- or length-for-age Z-scores among children aged less than 5y is ≥ 20% (de Benoist et al., 2007. The prevalence of low height- or length-for-age Z-scores (HAZ) for children 0–5y can be calculated from the WHO Child Growth Standard (WHO , 2006) and the computer program WHO (AnthroPlus). Note that in a healthy population of children the mean Z-score will be about 0.0 and the SD of the Z-score about 1.0; 2.5% of all the childen will have an HAZ-score < –2.

24c.4.2 Measurement of height or length

For infants and children ≤ 85cm (i.e. ≤ 2y), recumbent length is the recommended measure, preferably with the use of an infantometer with a range of 30–110cm, equipped with a digital counter reader. Recumbent length should be recorded to the nearest millimeter, or even more precisely (i.e., 0.1mm) when possible. Wooden or acrylic length measuring boards can be used, but they are rarely fitted with digital counters so are less reliable. Note that recumbent length for a child of about 2y is about 5mm greater than standing height for the same child (Haschke and van't Hof, 2000).

Children > 85cm and adults should be measured in the standing position, preferably using a free-standing stadiometer (range 65–206cm), again equipped with a digital counter reader capable of measuring stature to 0.1mm. Platform scales with movable measuring rods should not be used as they are less accurate. Clothing should be minimal when height measurements are taken so that posture can be clearly seen. Shoes and socks should not be worn. The timing of the measurement should be recorded; diurnal variations in height occur due to compression of the spine as the day progresses (Buckler, 1978). Consequently, in population studies, standing height should always be measured at the same time of day, preferably in the afternoon.

When measuring recumbent length or standing height, attempts should be made to minimize measurement errors. In longitudinal studies involving sequential measurements on the same group of individuals, one person should conduct all of the measurements throughout the study to eliminate between-examiner errors. This is especially critical when growth velocity is estimated; growth increments are generally small and are associated with two error terms, one for each measurement occasion. Recommendations of the minimal intervals necessary to provide reliable data on growth increments during infancy and early childhood are available (de Onis et al., 2004). In the WHO MGRS, the minimal interval recommended for reliable data on length measurements was every two weeks for infants from 2–6 weeks of age, monthly for ages 2–12 months, and bimonthly in the second year. During adolescence, increments measured over 6 months are the minimum interval recommended (WHO, 1997). For shorter intervals, the combined errors may be too large in relation to the expected mean increment.

In large regional surveys, several well-trained anthropometrists are often needed to rotate among the participants to reduce the effect of measurement bias. Regular standardization sessions to assess both within- and between-examiner reliability should be conducted throughout the data collection period to maintain the quality of the measurements and to identify and correct systematic errors in the measurements; details of the procedures used in the WHO Multicenter Growth Reference Study (MGRS) are given in de Onis et al. (2004). An anthropometric training video prepared for the WHO MGRS is also available on request from WHO.

The WHO MGRS recommends that the maximum allowable difference in length for acceptable precision between measurements by two anthropometrists is 7.0mm. (de Onis et al., 2004). Details of the measurement techniques and standardization protocols for both recumbent length and stature are given. Statistical methods exist for removing anthropometric measurement errors from cross-sectional anthropometric data; details are given in Ulijaszek and Lourie (1994).

This page titled 28.4: Linear growth (24c.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.5: Urine zinc (24c.5)

Daily excretion of zinc in the urine is typically 0.3–0.6mg. A systematic review and meta-analysis examining the response of urinary zinc excretion to changes in dietary intake was conducted by Lowe and co-workers (2009). They concluded that urinary zinc is a useful biomarker of increases in exposure to zinc supplements (i.e. > 15mg Zn/d) in adults with adequate zinc status at baseline. However, only when dietary zinc intakes are very low (< 1mg dietary Zn/d) is a decrease in urinary zinc excretion in adults apparent (Baer and King, 1984; Johnson et al., 1993; King et al., 2000). Comparable studies on infants and children have not been conducted.

Urinary zinc excretion is confounded by a variety of clinical conditions that induce hyperzincuria. These include sickle cell disease, alcoholism and liver disease, certain renal diseases and infections, injury, and burns. Starvation, strenuous physical exercise, diabetes, and trauma may also increase urinary zinc excretion (King et al., 2015). Hypertensive patients on long-term therapy with chlorothiazide also exhibit hyperzincuria(Prasad, 1983). Certain physiological conditions also influence urinary zinc excretion. Increases have been reported during pregnancy with concentrations returning to prepregnant levels after delivery (Fung et al., 1997). Whether changes occur during lactation is uncertain (King et al., 2015). Consequently, the overall health and diet of an individual must always be evaluated when using urinary zinc as a biomarker of zinc exposure.

The BOND Zinc Expert Panel classified urinary zinc as a “potential” biomarker that shows promise (King et al., 2015), based on its usefulness as a marker of compliance with zinc supplementation programs when at least 15mg Zn/d was given (Lowe et al., 2009). They emphasize, however, that the need to collect a 24-hour urine sample and the lack of established cutoffs limit the usefulness of urinary zinc as a biomarker. Moreover, because urinary zinc only responds to very low dietary zinc intakes, its use as a biomarker of dietary zinc intake among free-living populations is limited.

24c.5.1 Measurement of urinary zinc excretion

The amounts of zinc excreted in the urine usually range from 300 to 600µg/d. A 24-h urine collection is preferred because diurnal variation in urinary zinc excretion may be significant. Measurements of urinary zinc are usually performed by flame AAS.

This page titled 28.5: Urine zinc (24c.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.6: Hair zinc concentrations (24c.6)

Zinc is incorporated into the hair matrix when the hair is exposed to the blood supply during synthesis within the dermal papilla (Kempson et al., 2007). When the growing hair approaches the skin surface, the hair undergoes keratinization and the zinc accumulated during its formation becomes sealed into keratin protein structures and isolated from metabolic processes. Hence, the zinc content of the hair shaft reflects the quantity of zinc available in the blood supply at the time of the hair growth, not at the time of sampling (Kempson et al., 2007). Consequently, positive correlations between hair and serum zinc concentrations can only be expected in settings where zinc status is unchanged or chronic zinc deficiency exists. Assuming a normal rate of hair growth (i.e., about 1cm hair growth/month), the zinc concentration in the proximal 1–2cm of hair (i.e., the hair closest to the scalp) reflects the zinc uptake by the follicle 4–8 weeks before sample collection (Hopps, 1977). In cases where hair growth is arrested, as may occur in severe acute malnutrition (Erten et al., 1978) and acrodermatitis enteropathica (Bradfield and Hambidge, 1980), hair zinc should not be used. In such cases, hair zinc concentrations may be normal or even high. The advantages of using hair as a biomarker are listed in Box 24c.6.

Box 24c.6 The advantages of using hair as a biomarker. From IZiNCG Technical Brief #8. (IZiNCG, 2018).

- Concentrations are higher than in serum or urine making their measurement easier.
- Specimen collection is relatively non-invasive and samples can be collected and stored at room temperature without deterioration and the need for speial prservatives.
- Concentrations are not subject to the rapid fluctuations produced by a recent meal, diurnal or circadian variation, or infection, as seen in serum zinc,
- Concentrations reflect exposure over a longer retrospective time frame than serum or urine.

Nevertheless, several technical and biological factors may affect hair zinc concentrations. Hair is exposed to exogenous surface contamination so all specimens must be washed using a recommended procedure. Biological factors may include age, possibly sex, season of the year, and rate of growth, as noted earlier (King et al., 2015; IZiNCG, 2018). Neither cosmetic treatments or hair color affect hair zinc concentrations, provided appropriate washing methods are used (Kempson et al., 2007). The effects of all these possible confounding factors must be considered when interpreting hair zinc concentrations. The BOND Zinc Expert Panel classified hair zinc as a “potential” biomarker that shows promise, but requires the establishment of specific cutoffs to indicate zinc inadequacy in populations (King et al., 2015).

In part, the use of hair zinc concentration as a biomarker of zinc exposure has been hampered by the failure of commercial laboratories to adopt standardized methods for sampling and washing the hair samples, and to report the accuracy and precision of their analytical methods(Hambidge, 1982; King et al., 2015). Hair zinc concentrations were shown to respond positively and significantly to supplemental zinc in a systematic review and meta-analysis of three studies in healthy adults (Lowe et al., 2009), although their response to zinc depletion is uncertain(Lowe et al., 2009). In children, responses of hair zinc to supplemental zinc have been inconsistent, although some associations between low hair zinc concentrations and zinc-related functional outcomes (e.g., impaired taste acuity, appetite, linear growth, recurrent respiratory tract infection) have been reported (Hambidge et al., 1972; Buzina et al., 1980; Chen et al., 1985; Smit-Vanderkooy and Gibson, 1987; Gibson et al., 1989a; Cavan et al., 1993a; Ferguson et al., 1993). In addition, some studies have shown significant relationships between hair zinc concentrations and dietary zinc indices, most notably dietary phytate:zinc molar ratios, in individuals consuming predominantly plant-based diets (Ferguson et al., 1989; Gibson and Huddle, 1998).

24c.6.1 Interpretive criteria

There are no universally accepted reference values for hair zinc concentrations (Mikulewicz et al., 2013), and, as a result, the use of hairr zinc as a biomarker for assessing risk of zinc deficiency in populations has been limited. None of the reference values published to date have been compiled from a nationally representative reference sample of well nourished, healthy individuals free from conditions known to affect zinc status, unlike the IZiNCG procedure used to define reference values for serum zinc (King et al., 2015). In addition, standardized procedures for the sample collection, washing, and chemical analysis of the specimens, have not been employed (Mikulewicz et al., 2013). Hence there is an urgent need to compile a set of universal reference values for hair zinc concentrations.

Guidelines for establishing universal reference values are available in Mikulewicz et al. (2013). Reference values should be represented by the geometric mean ± CV (%) or median for hair zinc concentrations, preferably by sex and life-stage group. Reference limits indicative of unusually low hair zinc concentrations, represented by the 2.5th percentiles for males and females for each life-stage group, should also be determined. With these data, the feasibility of employing hair zinc to assess the likely risk of zinc deficiency in future national nutrition surveys could be explored.

Table 24c.7 presents the best available reference values for hair zinc (µg/g) derived from published studies for children ranging in age from 3 13y.Note data for both sexes are combined in five of the study groups in Table 24c.7, and that younger children tend to have lower hair zinc concentrations (IZiNCG, 2018, Table 24c.7).

Table 24c.7 Reference values for hair zinc (µg/g) based on healthy volunteer children from urban areas. From IZiNCG Technical Brie

Country Reference	Age (y)	Sex	n
Italy (Senofonte et al., 2000)	3–6	F/M	58
Korea (Park et al., 2007)	3–6	F/M	655
Italy (Senofonte et al., 2000)	6–10	F/M	96
Italy (Senofonte et al., 2000)	10–13	F/M	258
Italy (Dongarra et al., 2011)	11–13	F/M	130
Belgium (Vanaelst et al., 2012)	6–10	F	218

A cutoff point of 70µg/g (1.07µmol/g) has been established for hair zinc in young children that appears to indicate risk of zinc deficiency in both clinical and population studies. Cutoff points, unlike reference limits, are generally based on data from individuals with either clinical or functional manifestations of a nutrient deficiency. Zinc-related adverse health outcomes such as impairments in linear growth, appetite, and taste acuity (Hambidge et al., 1972; Smit-Vanderkooy and Gibson, 1987; Gibson et al., 1989a) have been associated with hair zinc concentrations less than 70µg/g (< 1.07µmol/g). A higher cut-off (< 110µg/g or < 1.68µmol/g) has been used in some studies for hair samples from children collected in the Autumn / Winter months to take into account the possible effect of season on hair zinc concentrations (Gibson et al., 1989b).

An additional application for hair zinc concentrations includes their use as a longer-term, retrospective measure of zinc exposure in case-control studies. Of note, is the existence of withon-person variability for hair zinc which has the potential to attenuate estimates of association in case-control or prospective cohort studies, as well as population prevalence estimates for risk of deficiency (Park et al., 2016). Such attenuation can be reduced by obtaining several replicate hair samples from each individual in the group and applying the mean value to represent true exposure.

24c.6.2 Measurement of hair zinc

Hair specimens should be collected from a representative sample of the target population or subgroups of interest during the same season of the year. Samples (at least 50mg) should be cut at skin level from the occipital region of the scalp (i.e., across the back of the head in a line between the top of the ears) with stainless steel scissors. Only the proximal (i.e., closest to the scalp) 1.0–2.0cm of the hair strands should be retained. These specimens will reflect the zinc uptake by the follicles 4–8 weeks prior to sample collection provided the rate of hair growth has been normal. Hair samples should be placed in labeled trace-element-free polyethylene bags for storage; any remaining hair strands should be discarded. Before washing the specimens to remove exogenous contaminants such as atmospheric pollutants, water, and sweat, any nits and lice should be removed if necessary using a microscope or magnifying glass and Teflon-coated tweezers. For each specimen, details of the ethnicity, age, gender, hair color, height, weight, date of collection, presence of malnutrition (where relevant), and use of dandruff shampoos or cosmetic treatments should always be recorded to aid in the interpretation of the data.

Washing with nonionic detergents (e.g., Triton X-100) with or without acetone is preferred for hair as these detergents are less likely to leach bound zinc from the hair and yet are effective in removing superficial adsorbed zinc. Washing with chelating agents such as EDTA should be avoided because they remove some of the tightly bound zinc that is an integral part of the hair sample (Shapcott, 1978). After washing and rinsing, the hair samples must be vacuum- or oven-dried depending on the chosen analytical method, and stored in a desiccator prior to laboratory analysis. For a detailed guide for measuring hair zinc concentrations the reader is advised to consult IZiNCG (Technical Brief No.8).

Several laboratory methods can be used to measure hair zinc concentrations in the washed specimens, depending on the instruments available. Washed specimens can be be prepared for analysis using microwave digestion when available, or wet or dry ashing, although the use of tetramethylammonium hydroxide (TMAH) to solubilize hair at room temperature warrants investigation. The latter method eliminates time-consuming ashing or wet digestion. Traditional analytical methods include flame AAS or multi-element ICP-MS, although non-destructive instrumental neutron activation analysis (INAA) can also be used. For INAA the washed hair specimens are placed in small, weighed trace-element-free polyethylene bags or tubes and oven-dried for 22 hours at 55°C. After cooling in a desiccator, the packaged specimens are sealed and weighed, prior to irradiation in a nuclear reactor.

Certified reference materials (CRM)s should always be used to assess the accuracy and precision of the chosen analytical method for hair zinc. Unsatisfactory quality control for hair trace element analysis among laboratories is a widespread problem (Hambidge, 1982; Mikulewicz et al., 2013). A CRM for human hair is available (e.g., Community Bureau of Reference, Certified Reference Material no. 397) from the Institute for Reference Materials and Measurements, Retieseweg, B-2440 Geel, Belgium. In-house controls prepared from a homogenous sample of finely cut digested hair (or powdered for INAA) should also be prepared and spiked with several different known quantities of zinc and the recoveries measured. The in-house controls should be analyzed in conjunction with a CRM periodically to monitor assay variations in the instrument and digestion procedures.

This page titled 28.6: Hair zinc concentrations (24c.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.7: Nail zinc (24c.7)

Fewer studies have examined nail zinc as a biomarker of zinc exposure than hair zinc even though the validity of nail selenium as a biomarker of selenium exposure in adults has been confirmed by several reports (Hunter et al., 1990). Nails, like hair, also incorporate zinc into their matrix when they are exposed to the blood supply within the germinal layer of the nail matrix, reflecting the quantity of zinc available in the blood supply at the time of the nail synthesis (He, 2011). The normal rate of growth for nails ranges from 1.6mm per month for toenails to about 3.5mm per month for fingernails (He, 2011). In cases where nail growth is arrested, as may occur in onychophagia (compulsive nail biting), nail zinc should not be used (He, 2011).

In healthy individuals not exposed to high levels of zinc in the atmosphere, nail zinc levels are similar to those in hair, with concentrations ranging from 80 to 200µg/g (1.22 to 3.06µmol/g). For example, in German school children aged 3–7years, the average toenail zinc concentration was about 130µg/g (1.99µmol/g) (Wilhelm et al., 1991), although nail zinc concentrations as high as 2000µg/g (30.6µmol/g)have been reported in galvanizers with a high exposure to zinc in the atmosphere (King et al., 2015). In a recent study of young children in Laos, nail zinc concentrations were higher at endline in those children receiving a daily preventive zinc supplement (7–10mg Zn/d) for 32–40 weeks compared to those being given only a therapeutic zinc dose (20g) for 10 days (geometric mean, 95%CI) (115.8, 111.6–119.9 vs. 110.4, 106–104.8µg/g.; p=0.055)(Wessells et al., 2014). Nail zinc concentrations can also be used as a longer-term retrospective measure of zinc exposure in case-control studies. For example, in a prospective study of US urban adults, toenail zinc was assessed in relation to the incidence of diabetes but no significant longitudinal association was found (Park et al., 2016).

Nails are exposed to exogenous surface contaminants so all specimens must always be washed using a recommended procedure. Chemicals introduced by nail polish can be removed by washing (He, 2011). Biological factors also affect nail zinc concentrations and include age, possibly sex, rate of growth, and onychophagia (compulsive nail biting).

The BOND Zinc Expert Panel classified nail zinc as an “emerging” biomarker with a theoretical association with zinc intake or status, although the sensitivity and specificity to any changes in zinc nutrition need further study (King et al., 2015). No universally accepted reference values exist for fingernail or toenail zinc concentrations exist and the reader is referred to Mikulewicz et al. (2013) for guidelines on their compilation.

24c.7.1 Measurement of nail zinc

Nails from all fingers or toes (at least 50mg) should be cleaned before clipping, and then stored in labeled T-E free polyethylene bags prior to washing. Bank et al. (1981) recommend using an aqueous nonionic detergent, followed by vacuum drying for traditional analytical techniques. For non-destructive analytical methods such as INAA and the newer technique involving laser-induced breakdown spectroscopy (LIBS), cleaning fingernail clippings with acetone (analytical grade) in an ultrasonic bath for 10 minutes followed by drying in air for 20–30 minutes is recommended (Riberdy et al., 2017). All dried specimens should be stored in a desiccator after weighing, prior to analysis. The wet and dry ashing methods and the analytical techniques outlined above for hair zinc concentrations can also be used for nail zinc. Alternatively, tetramethylammonium hydroxide (TMAH) can be used to solubilize nails at room temperature (Batista et al., 2008).

Laser-induced breakdown spectroscopy (LIBS) can be used and field-portable hand-held instruments operated from a battery have been developed. Investigations for measuring fingernail zinc *in situ* are underway (Riberdy et al., 2017). Preliminary results suggest that the *in situ*measurement of fingernail zinc by LIBS has potential as a non-invasive, convenient screening tool for identifying zinc deficiency in populations but may lack the precision required to generate absolute concentrations to characterize individuals (Riberdy et al., 2017). Research is ongoing to enhance the precision of this *in situ* method, and confirm its sensitivity and specificity to changes in zinc nutrition.

Unlike for hair, no Certified Reference Material exists for nail trace element analysis. Instead, in-house controls prepared from homogenous samples of powdered fingernails and toenails should be prepared, digested if necessary, and spiked with several different known quantities of zinc and the recoveries measured. Alternatively, an aliquot of the in-house control can be sent to an external quality laboratory and the results compared.

This page titled 28.7: Nail zinc (24c.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.8: Metallothionein concentrations in circulating blood cell types (24c.8)

Metallothionein is a small, cysteine-rich, metal-binding protein that readily binds to zinc and heavy metals. Up to seven zinc molecules can bind to one metallothionein molecule, that may serve as a small zinc reserve for cells. The highest concentrations of metallothionein are found in the liver, kidney, pancreas, and intestine. Concentrations in the intestine and pancreas respond to changes in dietary zinc intake, suggesting that metallothionein assists in the maintenance of zinc homeostasis in these tissues (King, 2011). Metallothionein also has a role in the acute phase response associated with inflammation when hepatic synthesis of metallothionein is activated, causing the redistribution of zinc from the plasma to the liver (Raiten et al., 2015).

Smaller quantities of metallothionein are also present in all tissue cells, and concentrations in serum, erythrocytes, and monocytes have been investigated as potential biomarkers of human zinc status. Serum metallothionein concentrations reflect changes in hepatic metallothionein, so any increase in hepatic levels in response to stress, inflammation etc., will result in concomitantly elevated levels in serum, thus limiting the use of serum metallothionein as a biomarker of zinc status.

In contrast to serum, erythrocyte metallothionein is not responsive to stress and appears to be sensitive to severe and moderate restrictions in dietary zinc intake, based on results of depletion-repletion studies (Grider et al., 1990; Thomas et al., 1992). However, because most of the metallothionein in erythrocytes is concentrated in the reticulocyte fraction, changes in erythropoiesis, as may occur with poor iron status, influence total erythrocyte metallothionein levels. With the advent of commercially available immunoassays for metallothionein, more research on the response of metallothionein concentrations in erythrocytes and other blood cell types to changes in dietary zinc intake or status in community-based studies is now possible. However, reference material for metallothionein concentrations is also needed to allow comparison of metallothionein concentrations using different immunoassays across studies (Hennigar et al., 2016). At present, the sensitivity and specificity of metallothionein concentrations in erythrocytes and other blood cell types and their usefulness in community-based studies remains to be established. As a consequence, metallothionein concentration in circulating blood cell types was classified as an “emerging” biomarker of zinc exposure and status by the BOND Zinc Expert Panel (King et al., 2015).

This page titled 28.8: Metallothionein concentrations in circulating blood cell types (24c.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)

Expression of the metallothionein gene is induced by zinc through the binding of zinc to metal-binding-regulatory transcription factor 1 (MTF-1). This transcription factor is known to be sensitive to cellular zinc concentrations, with metallothionein synthesis increasing as cellular zinc increases. Reverse transcriptase (RT) polymerase chain reaction (PCR) assays can be used to measure the response of metallothionein expression in various blood cell types to changes in dietary zinc. Hennigar et al. (2016), conducted a systematic review of sixteen experimentally controlled studies to determine the reliability and sensitivity of metallothionein expression to changes in dietary zinc in a variety of blood cell types. The studies were classified into three groups depending on the amount of zinc consumed per day (< 5, 15–22, and 50mg Zn/d). Only studies on healthy adults (> 18y) with no pre-existing health conditions, and that included baseline measurements or a placebo, were included. In all the studies, a depletion phase (mostly about 10 days) followed by repletion or a supplementation phase (ranging from 10 to 84 days) were included, although some also had a post-supplementation phase during which participants received a placebo tablet. Figure 24c.8

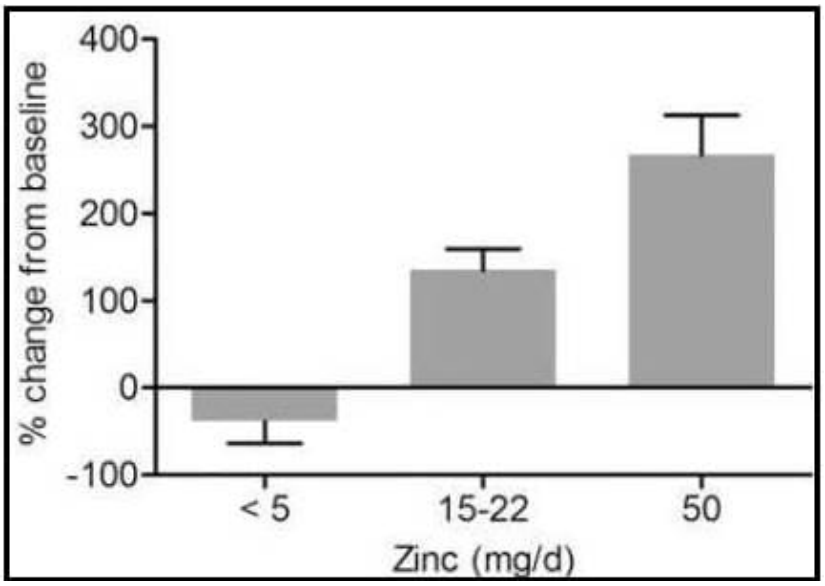


Figure 24c.8 Metallothionein expression in leukocytes. Values are percentage changes from baseline±SEs for all studies that examined metallothionein expression in leukocytes. In studies that determined metallothionein expression at multiple time points(19–21), the day with the greatest percentage change was used. Data are grouped based on the amount of dietary zinc consumed per day. From Hennigar et.al. (2016).

presents the estimated effect of varying concentrations of supplemental zinc on metallothionein expression in leukocytes based on the percentage change from baseline. There was a 39% decrease from baseline in metallothionein expression in leukocytes in participants consuming < 5mg Zn/d, and increases from baseline of 135% and 267% for those receiving 15–22mg Zn/d and 50mg Zn/d, respectively. Hence these findings indicate that metallothionein expression in leukocytes is sensitive to changes in dietary zinc intake, decreasing in zinc depletion, and increasing in response to zinc supplementation in a dose-dependent manner. However, these changes were not associated with consistent changes in plasma zinc, which suggests that metallothionein expression in leukocytes may be a more sensitive biomarker than plasma zinc for determining zinc exposure. Of interest was the lack of sensitivity of metallothionein expression in erythrocytes to changes in dietary zinc.

Clearly, further work is needed to investigate whether metallothionein expression in leukocytes can be used as a biomarker of zinc status among other age groups and free-living volunteers. The effect of potential confounders (e.g., inflammatory agents, free radicals, glucocorticoids, and pharmacologic agents)needs to be explored. In addition, more research is needed on whether the sensitivity to zinc supplementation and depletion is affected by variations in the proportions of blood cell subtypes among individuals. The abundance of metallothionein transcript in monocytes is three times that of granulocytes, and twice that found in

T lymphocytes (Aydemir et al., 2006). In view of these uncertainties, metallothionein expression in various blood cell types was classified as an “emerging biomarker” by the BOND Zinc Expert Panel (King et al., 2015).

This page titled 28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)

The discovery of shifts in the expression of the two families of zinc transporters: ZIP 1-14 (zinc “importers”) and ZnT 1-10 (“zinc exporters”) with changes in cellular zinc has stimulated investigations on the usefulness of zinc transporter mRNA expression in circulating blood cell types as biomarkers of zinc status (King et al., 2015). Hennigar et al. (2016), in their systematic review also investigated the response of the expression of several ZIP and ZnT zinc transporter proteins to zinc supplementation and depletion in various blood cell types isolated from healthy adults. Of the transporters, ZIP1 and ZnT1 were the most commonly measured proteins in the review. Inconsistent changes in the zinc transporter expression levels in response to zinc supplementation and depletion across studies were reported. Consequently, more studies are needed to assess the usefulness of zinc transporter expression in circulating blood cell types as a biomarker of zinc exposure or status, especially under free-living conditions when a more typical range of zinc intakes is consumed. (Hennigar et al. 2016) Selection of peripheral blood mononuclear cells as the tissue source for such investigations may be an advantage given the suggestion that immune calls may be the first to respond to a change in zinc status even before plasma zinc falls below the normal range (Prasad, 1998).

This page titled 28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.11: Oxidative stress and DNA integrity (24c.11)

Oxidative stress is caused by an imbalance between the increased production of multiple reactive oxygen species (ROS) (superoxide, hydrogen peroxide, and hydroxyl radicals) and the decreased protective action of antioxidants that are responsible for the neutralization and removal of ROS. Oxidative stress arising from this imbalance has the potential to damage proteins, deoxyribonucleic acid (DNA), and lipids, and may serve as a potential predictor in the development of different ROS-dependent diseases (e.g.,vascular diseases). Zinc plays an important role in antioxidant defence and the maintenance of cellular DNA integrity. As zinc itself is redox inert, it does not function as an antioxidant *per se,*instead functioning indirectly as an antioxidant and as a "pro-oxidant" over a limited range of zinc concentrations. Consequently, the term " pro-antioxidant" is used to describe the indirect functions of zinc as an antioxidant. In contrast, outside this range, zinc is a pro-oxidant. Hence, an increase in cellular oxidative stress and DNA damage may be an early sign of both reductions in cellular zinc as well as zinc overload. For more details see Kloubert and Rink (2015).

Several human studies have investigated whether intracellular DNA strand breaks, assayed using the comet assay, could serve as a functional biomarker of zinc status. In an experimentally-controlled depletion-repletion study in adult men with low intakes of dietary zinc for six weeks (i.e., 0.6mg Zn/d for one week followed by 4mg Zn/d for five weeks), the number of leukocytic DNA strand breaks increased, but declined when dietary zinc intakes were increased to 11mg Zn/d for four weeks (Song et al., 2009). Another study of adult men in which dietary zinc intakes containing 6mg Zn/d (with added phytate) for two weeks were increased to 10mg Zn/d for 4 weeks, also showed improvements in the repair of DNA strand breaks with increased intakes of dietary zinc. Serum protein concentrations associated with the DNA repair process also increased, despite no change in plasma zinc (Zyba et al., 2017). In a field setting in Ethiopia a decrease in DNA strand breaks was observed after supplementing women with 20mg zinc as zinc sulfate or placebo daily for 17 days, again despite no significant changes in plasma zinc.

Taken together, these findings indicate that even modest changes in dietary zinc appear to modulate DNA damage, primarily by reducing cellular oxidative stress, and confirm the sensitivity of the comet assay; see Singh et al. (1988) for assay details. Of note, however, despite modest increases in dietary zinc, there were no measured changes in plasma zinc concentrations, exchangeable zinc pool, erythrocyte or leukocyte zinc concentrations and metallothionein in leukocytes. These findings indicate that zinc attenuates oxidant stress over the range of usual zinc intakes, and in that capacity zinc modulates DNA damage. Nevertheless, many other conditions alter the redox state, so markers of oxidative stress, including DNA strand breaks, are not specific biomarkers for zinc nutrition. Hence, oxidative stress and DNA integrity were classified as "emerging" biomarkers by the BOND Zinc Expert Panel (King et al., 2015).

This page titled 28.11: Oxidative stress and DNA integrity (24c.11) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.12: Kinetic markers: pool sizes and turnover rates (24c.12)

Isotopic tracer studies in conjunction with a model-based compartmental analysis have identified and measured a relatively small whole-body exchangeable pool of zinc (EZP). This pool comprises the most metabolically active forms of zinc in the plasma, extracellular fluid, liver, pancreas, kidney, and intestine that make up about 10% of the whole-body zinc. The EZP provides zinc for zinc-dependent functions throughout the body and has a turnover rate of about 12.5 days. The size of EZP is estimated to contain 150–200mg zinc in adults, although varying markedly on a per kilogram body weight basis between adults and infants: 2.5mg Zn/kg in adults vs. 4.5mg Zn/kg in infants (King et al., 2001; Krebs et al., 2003).

Some but not all studies have shown a relation between dietary zinc and the size of the EZP. For example, the size of the EZP was significantly reduced in adults by severe dietary zinc restrictions (i.e, < 1mg Zn/d for 4–5 weeks) (King et al., 2001) or by chronically low habitual zinc intakes (i.e., average 5.2mg Zn/d) (Sian et al., 1996), but not by modest short-term changes in zinc intake (i.e., 4.6mg Zn/d for 10 weeks) (Pinna et al., 2001). Long-term (6 months) moderate zinc supplementation in male adults was shown to increase the size of EZP, which was reported to correlate positively with total zinc intake (Feillet-Coudray et al., 2005).

Of interest is the finding that during the study of severe restriction of dietary zinc (i.e., < 1mg Zn/d for 4–5 weeks) (King et al., 2001), the decline in EZP was nearly comparable to that of plasma zinc (i.e., 60% vs. 65%). This finding suggests that the EZP is part of the vulnerable pool susceptible to zinc depletion, although not more sensitive than plasma zinc concentrations. Whether the size of EZP is reduced by the metabolic redistribution of zinc due to inflammation and stress remains uncertain, although such a decrease is likely since plasma zinc is a component of the EZP (King, 2011). Other factors shown to influence the size of the EZP besides the level of dietary zinc include zinc absorption, age, and sex, with men having a larger pool size than women (Pinna et al., 2001; Krebs et al., 2003; Ruz et al., 2011).

Use of the EZP size as a biomarker of zinc nutrition remains limited, in part because of uncertainties about age and sex-dependent cutoffs for zinc insufficiency, the limited sensitivity of EZP to small changes in dietary zinc intake, and the relationship between EZP size and zinc function. In addition, the measurement of EZP necessitates the use of stable isotopes and a mass spectrometer for analysis (King, 2011; King et al., 2015).

Isotopic tracer studies have also been used to investigate changes in plasma zinc turnover rates as a biomarker of zinc status. When the total body zinc content is reduced, plasma zinc turnover rates are increased to meet tissue needs. Normally, total plasma zinc turns over about 150 times per day to provide zinc for numerous functions throughout the body. Increases in the plasma turnover rate from about 150–200 times per day have been reported in healthy men as a result of acute, severe zinc depletion (i.e., 0.23mg Zn/d) (King et al., 2001), but not when zinc depletion was more modest (4.6mg Zn/d for 10 weeks) (Pinna et al., 2001). In a study in premenopausal women (n=33), lower intakes of beef, a good source of absorbable zinc, were associated with a reduction in the plasma zinc turnover rate and a reduction in taste acuity (Yokoi et al., 2003). More research is required to establish the validity of plasma zinc turnover rate as a biomarker of zinc status and to standardize the conditions for making the measurements.

The BOND Zinc Expert Group classified the kinetic markers based on pool sizes and plasma zinc turnover rates as “emerging” biomarkers as they require more research to establish their sensitivity and specificity to changes in zinc nutrition (King et al., 2015; Allan et al., 2000).

This page titled 28.12: Kinetic markers: pool sizes and turnover rates (24c.12) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.13: Taste acuity (24c.13)

Diminished taste acuity (hypogeusia) is a non-specific characteristic of marginal zinc deficiency and has been investigated as a functional bioindicator of zinc status in both cross-sectional and experimentally controlled studies in children and adults (Hambidge et al., 1972a; Buzina et al., 1980; Gibson et al., 1989a; Wright et al., 1981). For example, in a study of boys 5–7y with low height percentiles and low hair zinc concentrations, impaired taste acuity was reported (Gibson et al., 1989a), but in a zinc depletion study of young men (0.25mg Zn/d for 4–9 weeks), there was a significant decline in their ability to discriminate differences in saltiness compared with baseline levels. This change was unrelated to changes in zinc concentrations in parotid saliva or plasma (Wright et al., 1981).

An age-related decline in taste acuity is well established in the elderly, but whether it is associated with a decrease in zinc status is uncertain. Some cross-sectional studies have reported positive associations between taste acuity for salt and erythrocyte zinc in the elderly (Stewart-Knox et al., 2005). Moreover, taste acuity for salt (but not for sweet, sour, or bitter taste) reportedly increased in older adults from Grenoble but not from Rome in response to 30mg Zn/d compared to placebo during a six-month double-blind randomized controlled trial (Stewart-Knox et al., 2008). Loss of taste during irradiation or chemotherapy has been restored with zinc supplementation in some (Yamagata et al., 2003; Najafizade et al., 2013), but not all (Halyard et al., 2007; Khan et al., 2019) studies.

Measurement of taste acuity

Several methods for testing taste acuity are available. Usually, evaluation is based on assessment of both the detection and recognition thresholds for each of four taste qualities (e.g., sweet, salt, sour, bitter), and sometimes umami (monosodium glutamate). The detection threshold is defined as “the lowest concentration at which a taste can just be detected”; the recognition threshold is “the lowest concentration at which the quality of the taste stimulus can be recognized”. The two thresholds are each determined for one taste quality before proceeding to the next. Several investigators have used this technique on young children (Hambidge et al., 1972a; Buzina et al., 1980).

An alternative technique that assesses only recognition thresholds and is better suited to children who are easily distracted and have short attention spans, has been developed (Desor and Maller, 1975). In some of these studies on children, recognition thresholds for only one taste quality (often salt) have been determined. Salt is selected because taste perception to moderately salty solutions changed significantly during a zinc-depletion study of young adult men (Wright et al., 1981). Recognition thresholds for salt have been measured in a study of Canadian school boys with low height percentiles and in Guatemalan school children. In these two investigations, school children with low hair zinc concentrations had higher recognition thresholds for salt than children with hair zinc concentrations indicative of “normal” zinc status (Gibson et al., 1989a; Cavan et al., 1993a). Sodium chloride concentrations used for testing recognition thresholds in these school children were 10, 15, 20, and 25mmol/L (Gibson et al., 1989a). Participants rinse a small amount (10mL) of each solution around in their mouths, expectorate it, and are then asked to identify the presented sample as salty or plain water. Salt solutions of increasing or decreasing concentrations are used, where appropriate, until the subjects correctly identify salt at one concentration and fail to do so at the next lower concentration. Only 2 out of 10 judgments are allowed to be incorrect before moving to the next higher or lower salt concentration. The midpoint between these two concentrations is used as the recognition threshold for salt for each participant.

Some studies have used an electrogustometer to measure taste thresholds by applying a weak electric current to the tongue (Prosser et al., 2010). A threshold value is determined by alternately lowering and raising the current todetermine the smallest stimulus that can be discriminated correctly (Grant et al., 1987). For a review of the inherent strengths and limitations of electrogustometry, see Stillman et al. (2003). Automated-computer-controlled electrogustometry can now be used to estimate taste detection thresholds (Stillman et al., 2000), although so far, this method has not been applied in studies of zinc status. For all taste acuity methods, it is preferable to perform the tests midmorning, at least 2 hours after a meal, with the same person administering the test on each occasion. Many other factors affect taste function. Hence, taste acuity should be used in conjunction with other biomarkers to assess zinc status. The BOND Zinc Expert Panel classified taste acuity as an “emerging” zinc biomarker.

This page titled 28.13: Taste acuity (24c.13) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.14: Zinc-dependent enzymes (24c.14)

Although zinc serves in a catalytic role for over three hundred zinc metalloenzymes, no single zinc-dependent enzyme has gained widespread acceptance as a biomarker of zinc status in humans. Their response to zinc deficiency, even in experimental studies of zinc depletion-repletion, has been very equivocal, and none have proved to be a consistent biomarker of zinc status. The physiological symptoms of zinc deficiency probably reflect, a series of biochemical changes (King, 2011).

Several zinc metallo-enzymes have been investigated as possible biomarkers of zinc status in plasma, erythrocytes, erythrocyte membranes or in specific cell types. Examples include d-amino-levulinic acid dehydratase, angiotensin-1-converting enzyme, α-D-mannosidase, extracellular superoxide dismutase, nucleoside phosphorylase, carbonic anhydrases, ecto purine 5′nucleotidase, and alkaline phosphatase. Of these zinc metallo-enzymes, the activity of alkaline phosphatase (EC 3.1.3.1) in serum, erythrocytes, or erythrocyte membranes has been most frequently studied. In a systematic review of biomarkers of zinc status that included seven different enzymes (Lowe et al., 2009). Six studies investigated the response of plasma alkaline phosphatase. However, no significant effect of zinc intakes on overall plasma alkaline phosphatase was found after combining the data from zinc depletion and supplementation trials.

Alkaline phosphatase in the circulation consists of a mixture of three different isozymes (intestinal, placental, and liver/kidney/bone) and several isoforms. It is possible that assessment of their individual responses to changes in zinc intakes may provide more sensitive and consistent data (King et al., 2015). For the other six enzymes included in the systematic review, the number of studies was too limited to assess their effectiveness as biomarkers of zinc status (Lowe et al., 2009). Hence, based on the available evidence the BOND Zinc Expert Panel classified the activity of zinc-dependent enzymes as “not useful” as a biomarker of zinc exposure or status (King et al., 2015).

This page titled 28.14: Zinc-dependent enzymes (24c.14) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.15: Erythrocyte zinc (24c.15)

The concentration of zinc in erythrocytes is approximately ten times higher than in plasma, and normally 8–14μg/g wet weight or10–11μg/10^{10}cells. Zinc is present mainly as carbonic anhydrase (EC 4.2.1.1) (80–88%), with a small amount bound to Cu,Zn-SOD (EC 1.15.1.1) (~5%). Approximately 2%–3% is bound to low-molecular-weight-binding ligands. The life span of erythrocytes is 120d.

The BOND Zinc Expert Panel classified erythrocyte zinc or erythrocyte membrane zinc as “not useful” based on the inconsistent response to changes in zinc intakes or plasma zinc concentrations (King et al., 2015). For example, in a 49d zinc depletion study, no changes in erythrocyte zinc were seen, despite evidence of impaired taste acuity and immune function, two potential bio-indicators of functional zinc depletion (Ruz et al., 1992). In addition, no response was reported in zinc supplementation studies of 30d and 90d duration, even when a dose of 50mg of zinc/day was used (Thomas et al., 1992; Davis et al., 2000). Taken together, the results emphasize the low sensitivity of erythrocyte zinc.

There are no standardized units for the expression of zinc concentrations in erythrocytes, making comparisons among studies difficult. The units can be based on hemoglobin, volume of packed cells, and per cell. There are presently no accepted standards for conversion among these units. No interpretive criteria have been established for erythrocyte zinc.

This page titled 28.15: Erythrocyte zinc (24c.15) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.16: Leukocyte zinc (24c.16)

Leukocytes contain up to twenty-five times more zinc than erythrocytes, and have a very much shorter lifespan than the 120-day lifespan of erythrocytes. Concentrations of zinc in leukocytes vary according to the cell type, with monocytes and lymphocytes having the highest concentrations, followed by neutrophils. Mixed white blood cells generally contain 75µg zinc/10^{10} cells.

The BOND Zinc Expert Panel classified concentrations of zinc in leukocytes and leukocyte subpopulations as “not useful” as biomarkers of zinc nutrition in view of the inconsistent results observed in studies of both experimental zinc depletion – repletion (Ruz et al., 1992) and prolonged high zinc supplementation (e.g., 50mg Zn/d) (King et al., 2015).

The separation of leukocytes and other specific cell types is difficult, and some of the inconsistencies reported for leukocyte zinc in zinc depletion-repletion studies may result from analyzing mixtures of cell types (e.g., neutrophils and lymphocytes), with different half-lives and zinc concentrations. Further, leukocytes and other specific cellular types are prone to contamination from exogenous sources of zinc and from platelets. Data are also difficult to interpret because no consensus exists for the method used to express zinc concentrations in leukocytes or specific cell types. Results can be expressed on a mg/kg dry weight basis to eliminate the effects of variable cell water content, but results expressed as ng/10^{6} cells are generally preferred (Nishi, 1980; Milne et al., 1985). Additional factors limiting the use of leukocyte zinc or the zinc content of subpopulations is the relatively large volumes of blood (i.e., about 5mL) required for the analysis, and the lack of established interpretive criteria (King et al., 2015).

This page titled 28.16: Leukocyte zinc (24c.16) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.17: Salivary zinc (24c.17)

Concentrations of zinc in saliva have been investigated as a measure of zinc status because zinc appears to be a component of gustin, an essential protein involved in taste acuity, that may be impaired by marginal zinc deficiency (Henkin, 1984). Zinc concentrations in mixed saliva, parotid saliva, salivary sediment, and salivary supernatant have all been investigated, but their use as biomarkers of zinc status is equivocal(Greger and Sickles, 1979; Freeland-Graves et al., 1981; Baer and King, 1984). As well, although the collection of saliva samples is quick and relatively noninvasive, the rate of flow and stimulation of saliva is difficult to control, and diurnal variation in zinc levels may occur (Warren et al., 1981). In view of these problems, and the contradictory results observed, the use of salivary zinc as a biomarker of zinc status has been abandoned.

This page titled 28.17: Salivary zinc (24c.17) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

28.18: Multiple biomarkers (24c.18)

The efficient regulation of zinc homeostasis complicates the assessment of zinc deficiency and excess, and currently there is no single, sensitive, specific biomarker of zinc status. At the population level, three biomarkers have been recommended by WHO / UNICEF / IAEA /IZiNCG (de Benoist et al., 2007) and endorsed by the BOND Zinc Expert Panel (King et al., 2015) to assess zinc status and/or evaluate nutrition interventions to combat zinc deficiency. They are itemized in Section 24c.1.7. Details on the measurement and interpretation of these biomarkers are available from IZiNCG (Technical Briefs No.2 and No.3) and a table listing the relative strengths and weaknesses of the biomarkers is summarized in (King et al., 2015). For each biomarker, an indicator has been developed to estimate risk of zinc deficiency at the population level. These are: (a) the prevalence of usual zinc intakes below the estimated average requirement (EAR), (b) the percentage of the population with low serum zinc concentrations below the appropriate reference limit, and (c) the percentage of children less than 5y with length- or height-for-age less than –2SD below the age-specific median of the reference population (WHO, 2006). For each indicator, a "trigger level" for the prevalence considered indicative of elevated risk and of public health concern is given, at which level an intervention to improve population zinc status is warranted. Ideally, all three indicators should be included in a survey to strengthen the conclusion regarding the zinc status of the population and the need for an intervention, although in practice, this is not always possible. de Benoist et al., 2007 suggest selecting indicators according to the level of available resources (Box 24c.7).

Box 24c.7. Resource availability guidelines

- **Limited resources available.** Review existing evidence: Stunting rates (height– or length–for–age Z-score < –2SD) of children under 5y; Dietary patterns and likelihood of providing adequate bioavailable zinc in the food supply; Results from clinical, community-based studies of functional response to zinc supplements; Results from community-based studies on dietary zinc intakes or serum zinc concentration
- **Some resources available** Augment existing information with data collected in a prospective survey. This would normally be linked with other health/nutrition surveys with broader goals: In at least a subsample of the survey population, or samples focused among subpopulations that are likely beneficiaries of nutrition or health interventions. Include serum zinc concentration and/or dietary zinc intake adequacy as biomarkers; where food-based interventions may be considered, at least 1 day of 24-hour recall per individual in the sample should be completed
- **Adequate resources available** Where resources permit: Serum zinc concentration should be determined in a representative sample of all relevant subpopulation groups; 24-hour dietary recalls, or other locally validated quantitative dietary assessment methods, should be repeated on nonconsecutive days on at least a subsample to estimate the distribution of usual zinc intakes

For individuals, unless zinc deficiency is severe, moderate or mild zinc deficiency can be present in the absence of abnormal zinc biomarkers, including serum zinc, despite the presence of disturbances such as growth retardation, loss of appetite, and impaired immune function. Hence, for assessing the zinc status of individuals, the recommended approach is based on a medical history, a qualitative assessment of the diet, and a clinical assessment, with plasma or serum zinc providing only supplementary data (King et al., 2015), as noted earlier.

This page titled 28.18: Multiple biomarkers (24c.18) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29: SELENIUM (CHAPTER 25B)

Abstract

In only a few decades, the trace element selenium (Se) was added to the Nutrition agenda as an essential nutrient with potential to reduce cancer risk. Found in the 1950's to prevent vitamin E deficiency disorders in animals, initial investigations in animal nutrition spawned studies of metabolic mechanisms starting in the 1960's. It is now known that Se functions as an essential constituent of some 25 proteins that contain a previously unrecognized Se-containing amino acid, selenocysteine. These selenoproteins are thought to discharge the nutritional functions of Se. The relationship of one, glutathione peroxidase, to the level of Se intake was used in 2000 to establish an RDA for Se. Se is typically obtained from protein-rich foods, particularly those produced on Se-rich soils. Low Se status occurs in areas of low soil Se, and among individuals consuming low-protein diets. Many consequences of Se deprivation are sub-clinical in nature, requiring other precipitating factors (vitamin E deficiency, low iodine intake, viruses, carcinogens) to reveal the effects of sub-optimal expression of selenoenzymes and/or insufficient amounts of active Se-metabolites. Studies that started in the 1960's showed that Se can prevent carcinogenesis in animal and cell models, and that anti-tumorigenic activities required Se doses greater than those required to support selenoenzyme expression. Only a few clinical intervention trials have been conducted to test the hypothesis that Se supplementation can reduce cancer risk in humans; most, but not all have shown positive results. Hence, whether Se may be a risk modifier for cancer remains unclear.

29.1: Overview (25b.1)

29.2: Selenium Chemistry (25b.2)

29.3: Health Effects (25b.3)

29.4: Selenium Requirements (25b.4)

29.5: Food Sources (25b.5)

29.6: Enteric Absorption (25b.6)

29.7: Metabolism (25b.7)

29.8: Biochemical Functions (25b.8)

29.9: Assessing Selenium Status (25b.9)

29.10: Questions Remain (25b.10)

This page titled 29: Selenium (Chapter 25b) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.1: Overview (25b.1)

Selenium (Se) is an essential nutrient with several multiple biological activities. It was first recognized for its toxic potential in the 1930's when it Se was identified as an active principal in a plant-induced neuropathy of grazing horses and cattle in the northern Great Plains of the United States (Painter, 1941). In the 1950s its nutritional role was discovered with experiments with laboratory and livestock species that found the element to prevent pathologies in vitamin E-deficient animals (Schwarz & Foltz, 1957; Schwarz et al.,1957; Patterson et al., 1957). Then, in the 1980s its role in human health was indicated with the finding that very low blood Se levels were associated with an endemic heart disease in women and children in parts of China (Ge et al., 1983), and Se-supplementation was found to be preventive by the Keshan Disease Research Group (1979). In parallel studies starting in the 1960s, Se was shown to have anti-tumorigenic capabilities in a variety of animal models (Combs and Lu, 2006; Yan and DeMars, 2012; Yan and Combs, 2014). Clinical trials later showed cancer risk reduction for subjects of low-to-moderate Se status (Yu et al., 1991; Blot et al., 1993; Taylor et al., 1994; Blot et al., 1995; Clark et al., 1996; Yu et al., 1997), ,although one showed no effect (Lippman et al., 2009).

That the various roles of Se have different dose-response relationships has been demonstrated in experimental animals. Biochemical and physiological lesions of Se deficiency are prevented by relatively low dietary levels, e.g., 0.1mg/kg (Combs and Combs, 1986). Reductions in experimental tumorigenesis occur when Se is fed at 10–20 times the nutritional level, i.e., ≥ 1.5mg/kg (Ip and Ganther, 1990; Ip, 1998). Adverse effects are not seen unless still greater dietary levels are used, i.e., > 5mg/kg (Combs and Combs, 1986). Because Se intakes of free-living people can seldom be ascertained accurately, Se status in humans must be inferred from measures related to the intake and disposition of Se.

This page titled 29.1: Overview (25b.1) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.2: Selenium Chemistry (25b.2)

The chemistry of Se has similarities to that of sulfur (Combs and Combs, 1986; Combs, 2016). Hence, Se can be incorporated in plants and animal tissues in lieu of sulfur. However, Se-compounds have higher redox potentials. Thus, while sulfides are metabolized to sulfates, selenate is metabolized to selenide. The biochemistry of Se involves mainly its three anionic forms: Se^{2-} (selenide and hydrogen selenide [H_2Se]), SeO2+
3(selenite and selenous acid (H_2SeO_3), and SeO2+

4 (selenite and selenic acid [H_2SeO_4]). Selenide is pivotal in Se metabolism, as HSe^- is the obligate precursor for the biosynthesis of selenocysteine (SeCys) , the functional form of the element. Selenite is easily reduced. At low pH it is readily reduced by such agents such as ascorbic acid and SO_2. It can also readily react with nonprotein thiols (e.g., reduced glutathione [GSH]) or with protein-sulfhydryl groups subsequently to be reduced to selenide via sequential formation of selenotrisulfides (e.g., oxidized glutathione [GSSeSG]) and selenodisulfides (e.g., glutathione selenol [GSSeH]). Selenate is much more resistant to reduction; it is activated to adenosine-5-phosphoselenate as a first step in its physiological reduction.

The major forms of Se in cells are the selenoamino acids selenocysteine and selenomethionine (SeMet). In each, Se is covalently bound to carbon. selenocysteine is co-translationally synthesized in animal tissues; whereas, selenomethionine is produced in plants and microorganisms using elemental Se in lieu sulfur. Animal tissues can incorporate selenomethionine into proteins nonspecifically in place of methionine (Met).

This page titled 29.2: Selenium Chemistry (25b.2) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.3: Health Effects (25b.3)

25b.3.1 Animals

Selenium was recognized as an essential nutrient when it was found to be the active component of yeast that prevented necrotic liver degeneration in vitamin E-deficient rats and exudative diathesis in vitamin E-deficient chicks (Schwarz and Foltz, 1957; Schwarz et al., 1957; Patterson et al., 1957). Se deficiency is now known to play roles, with vitamin E, in preventing several nutritional diseases of livestock (Combs and Combs, 1986; Surai and Fisinin, 2016): skeletal myopathy of lambs, calves, horses, chickens and ducks; gizzard myopathy in turkey poults; cardiomyopathy in pigs; placental retention in dairy cows; infertility in boars. Accordingly, Se, typically in the form of sodium selenite (Na_2SeO_3), is commonly included as a nutritional supplement in formulated animal feeds.

25b.3.2 Humans

Severe endemic Se deficiency has been linked with two diseases of humans: a juvenile cardiomyopathy (Keshan disease) and an osteoarthropathy (Kashin-Beck disease) (Combs and Combs, 1986). These disorders were described in rural, mountainous areas of central and northeastern China and Russia (eastern Siberia) where food systems were exceedingly low in Se (soils < 125µgSe/kg; grains < 40µgSe/kg) and humans had blood Se levels < 25ng/mL (compared to 85–200ng/mL in the United States).

Keshan disease is a multifocal myocarditis primarily affecting children 2–10y and, to a lesser extent, women of child-bearing age. Diagnosis is based on signs of acute or chronic cardiac insufficiency, cardiac enlargement, arrhythmia, and electrocardiographic abnormalities (Ge et al., 1983). Affected individuals may show cardiogenic shock or congestive heart failure. In the 1970s, Chinese scientists found Se effective in preventing the disease (Keshan Disease Research Group, 1979). This led to the widespread use of Se supplements (oral doses of sodium selenite at 0.5–1mg Se/wk, or table salt fortified with selenite at 10–15mg Se per kilogram body weight) (Keshan Disease Research Group, 1979; Chen et al., 1980). Such interventions have virtually eliminated Keshan disease from previously affected areas. It appears likely that Keshan disease may be caused by cardiophilic RNA viruses the virulence of which can be potentiated by severe Se deficiency through increased viral mutation (Beck and Levander, 1998; Beck, 2007). This suggests that Se deficiency may also increase risks for other diseases caused by RNA viruses (e.g., measles, influenza, hepatitis, and HIV/AIDS).

Kashin-Beck Disease is an osteoarthropathy affecting the epiphyseal and articular cartilage, and epiphyseal growth plates of growing bones. However, the etiology is unclear. It presents as enlarged joints (especially of fingers, toes and knees); shortened fingers, toes and extremities; and, in severe cases, dwarfism. A meta-analysis of 15 clinical trials (Zou et al., 2009) suggested that Se supplementation had some value in preventing the disease. However, additional factors, including iodine deficiency and exposure to fungal toxins, have been suggested as etiological factors.

Microbiome. Dietary Se may also affect the hindgut microbiome, which is involved in immunologic signaling, energy abstraction from non-digested carbohydrates, and carcinogenic risk. Se reaches the hindgut in the form of non-digested Se from the diet and endogenous Se from enterohepatic circulation and sloughed mucosal cells. Selenium has been shown to promote taxonomic diversity of the microbiome of the rat(Kasaikina et al., 2011), which is capable of producing methylated Se-metabolites independent of the host (Krittaphol et al., 2011). Because methanogenic *Archaea* and gram$^+$ bacteria express selenoproteins (Kryukov and Gladyshev, 2004; Foster, 2005). it is possible that these species may be growth-limited by Se supply.

25b.3.3 Anti-carcinogenesis

That Se may be anti-carcinogenic was first suggested in the late 1960s based on inverse relationships of cancer mortality rates and forage crop Se contents in the US (Shamberger and Frost, 1969; Shamberger and Willis, 1971).

Animal models. The results of virtually all animal and cell models used, show that both inorganic and organic Se compounds can prevent or reduce carcinogenesis (Ip, 1998). Some of these effects may involve actions of selenoenzymes preventing DNA damage by reducing hydroperoxides. Other anticarcinogenic effects, particularly those associated with supranutritional Se intakes, may involve Se metabolites increasing programmed cell death by way of redox cycling, attacking protein thiols and/or labilizing proteins critical for tumor cell growth and their spread (Jackson and Combs, 2011; Yan and DeMars, 2012; Combs and

Yan, 2016). The results of studies with animal tumor models are clear: Se can inhibit both primary and secondary carcinogenesis.

Studies by Ip and colleagues (Ip et al., 1991; Ip, 1998) indicated methylselenol (CH_3SeH, produced by the methylation of H_2Se; Fig 1) as active in Se-antitumorigenesis. CH_3SeH can modulate the expression of key genes in the p53 pathway and lymphocyte cell surface ligands that trigger immune activation Zeng et al., 2013). and reduces plasma proteases and angiogenic factors required for tumor spreading (Li et al., 2008; Yan et al., 2012; 2014). The involvement of CH_3SeH and perhaps other methylated Se-metabolites in Se-antitumorigenesis suggests underlying mechanisms that may become significant at supranutritional Se intakes (Jackson and Combs, 2008): redox cycling of H_2Se and/or CH_3SeH to generate reactive oxygen species (ROS); modifications of protein-thiols by H_2Se and/or CH_3SeH to affect protein function/turnover; Met mimicry by selenomethionine in proteins to sensitize critical regulatory proteins to ROS.

Humans. Whether Se supplementation can be useful in reducing cancer risk in human populations is less clear. Only a few clinical trials have been conducted to test the hypothesis that Se-supplementation can reduce cancer risk (Yu et al., 1991; Blot et al., 1993; Taylor et al., 1994; Blot et al., 1995; Clark et al., 1996; Yu et al., 1997; Lippman et al., 2009). While one trial with males of relatively high Se status found no effects on prostate cancer risk (Lippman et al., 2009); most found cancer risk reductions for subjects of low-to-moderate baseline Se status. One well designed trial found supplemental Se to reduce prostate cancer risks in subjects with baseline plasma Se < 106ng/mL, but not in those of greater baseline Se status (Combs et al., 2001; Duffield-Lillico et al., 2003). Meta-analyses of prospective studies and Se-intervention trials (Lee et al., 2011; Dennert et al., 2011) concluded that Se supplementation had a preventive effect on cancer incidence in populations with baseline serum Se levels < 126ng/mL and in high-risk populations for cancer. Subsequent systematic reviews (Vinceti et al., 2014; Cai et al., 2016) indicated that people of higher Se status or with greater Se intakes had lower risks of several types of cancer; but no protective effects of Se supplementation were detected (Cai et al., 2016). Overall, it appears that high Se status can reduce the risk of certain cancers in many people, especially those with low or nutritionally adequate, but not high, Se status (Combs and Yan, 2016).

25b.3.4 Selenosis

That Se can be toxic was recognized before its nutritional function. Almost nine decades ago it was found to be the an active principal in a neuropathy of grazing horses and cattle, grazing on Se-accumulator plants growing on seleniferous soils of the northern Great Plains of the United States (Painter, 1941). Subsequent investigations have characterized selenosis in both its acute and chronic forms (Combs and Combs, 1986).

Acute Selenosis has occurred in humans due to accidental ingestion of gram quantities of Se in high-Se solutions including gun bluing, sheep drench and anti-dandruff shampoo (Combs and Combs, 1986; Institute of Medicine, 2000). Each case involved rapid development of severe gastrointestinal and neurological symptoms followed by acute respiratory failure, myocardial infarction and renal failure. Some cases showed garlic odor of the breath, an early sign in Se-intoxicated animals (Franke and Moxon, 1938; Jiang et al., 1983) due to the excretion of dimethylselenide ($[CH_3]_2Se$) across the lung (Buchan, 1947; Skorniakova et al., 1969). In rodents, injected selenite causes a dose-dependent increase in excreted $(CH_3)_2Se$ (Tiwary et al., 2005) even at sub-toxic doses when it can represent nearly 30% of the dose (Yonemoto et al., 1985; Tandon, et al., 1986; Magos et al., 1987; Vandhanavikit et al., 1987).

Chronic Selenosis. Endemic selenosis occurred in the early 1960s in Enshi County, Hubei Province, China where residents consumed foods containing very high levels of Se (several hundred mg/kg) due to the use of the ash from local high-Se coal ash to amend agricultural soils (Yang et al., 1983, 1989b, 1989c). During the years of peak prevalence (1961-1964), the five most heavily affected villages experienced morbidity rates approaching 50% and some individuals having blood Se levels as high as 3,200ng/mL. The most common signs were hair and nail loss; lesions of the skin, nervous system and teeth were also reported. Dietary Se intakes, estimated a few years later, averaged about 5mg/day, with individuals with Se intakes less 1500µg/day not showing adverse effects.

Diabetes Risk. Two clinical trials found plasma Se concentrations exceeding about140ng/mL to be associated with increased risk of type 2 diabetes (T2D) (Bleys et al., 2007; Stranges et al., 2007). Those findings were based on self-reports of T2D and/or medication use consistent with T2D, unconfirmed with measures of circulating glucose, insulin or glycated hemoglobin (HbA1c). They have not been confirmed in other randomized Se-intervention trials, including one in which subjects showed plasma Se concentrations exceeding 200ng/mL (Thomson et al., 1982). Nevertheless, a study of gestating rats found the high-level Se exposure (3mg Se/kg diet) of dams to induce insulin resistance and glucose intolerance late in gestation and in offspring after several weeks postpartum (Zeng et al., 2012). A similar experiment with pigs found fed a diet containing 3mg Se/kg to support greater circulating insulin levels, with normal glucose levels, compared to those fed the nutritional level(Liu et al., 2012).

This page titled 29.3: Health Effects (25b.3) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.4: Selenium Requirements (25b.4)

The Institute of Medicine of the National Academy of Sciences Recommended Dietary Allowance (RDA) was based on maximal expression of Glutathione Peroxidase 3 observed in two studies Institute of Medicine, 2000). The first study involved Chinese adults with a basal intake of 11µg Se/day supplemented with graded levels of selenite (Yang et al., 1987). It indicated that Glutathione Peroxidase 3 activity increased until it plateaued at a total Se intake of 41µg/day. Adjusting for differences in the body weights of the subjects and most Americans, this yielded an estimated requirement of 52µg Se/day. The second study involved adult New Zealanders with baseline intake of 28µg Se/day and supplemented with graded levels of Se as seleno-methionine (Duffield et al., 1999). Significant increases in Glutathione Peroxidase 3 activity were seen with 10µg supplemental Se/day, suggesting a requirement of 38µg Se/day. The IOM used the average of the two estimates, 45µg/day, as the Estimated Average Requirement (EAR). Assuming a 10% coefficient of variation, the RDA for was set at 55µg/day for both men and women. RDAs for children were extrapolated from this value based on body weight. Those for infants, however, were based on projected Se intakes from breast milk (birth–6mos: 15µg/day), and from breast milk plus complementary foods (6mos–1y: 20µg/day). Additional Se was recommended during pregnancy and lactation, bringing those RDAs to 60µg/day and 70µg/day, respectively. The World Health Organization (1996) based its recommendations on amounts of Se needed to support two-thirds of maximal Glutathione Peroxidase 3 activity. Based on the data of a single study (Yang et al., 1987), that intake was imputed to be 26µg Se/day. Adjusting for body weight and assumed interindividual variation, the WHO recommended Se intakes for men and women were set at 40 and 30µg Se/day, respectively. Neither of these groups considered the clear differences in the bioeffacies of selenite and the major form of Se in foods, selenomethionine. Xia et al. (2005) showed that selenomethionine was more effective than selenite in supporting Glutathione Peroxidase 3 expression in low-Se Chinese subjects with graded levels of Se over a 4-month period (Xia et al., 2005).

The US Food and Drug Administration set a “no observed adverse effect level” (NOAEL) for Se at a whole blood Se concentration of 1000ng/mL, which they calculated to correspond to a dietary intake of 853g/day in an adult male (Poirier, 1994). The Institute of Medicine Institute of Medicine, 2000). set a tolerable upper Se intake level (UL) for adults at 400ug/d, i.e., about half the NOAEL; other national bodies adopted similar recommendations, i.e., 350–450µg/d for adults (Hurst et al., 2013).

This page titled 29.4: Selenium Requirements (25b.4) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.5: Food Sources (25b.5)

Plant Foods. The Se contents of foods vary widely due to the differences in the soluble Se contents of soils, in the capacities for various plant species to access soil Se, and, for livestock products, in the amounts of biologically available Se in feeds. The Se contents of plant foods tend to be influenced by the local soil Se environment. For example, cruciferous vegetables, which are naturally rich in sulfur, can accumulate nutritionally significant amounts of Se (1–10mg/kg) if grown on high-Se soils or if fertilized with selenite or selenate salts. Studies in many countries have shown the effects of geographic variation in the Se contents of foods (Combs et al., 2001). Wheat grain may contain> 2–5mg Se/kg if produced in parts of the northern plains of North America; but as little as 0.1mg Se/kg if produced in Kansas or New Zealand; and only 0.005mg Se/kg if produced in Shaanxi Province, China. On a global basis, foods with the lowest Se contents are found in the provinces of Heilongjiang, northern Shaanxi, and Sichuan in China.

Animal Foods. Because of the intimate relationship between plants and animals in the food chain, the Se contents of foods of animal origins tend to be influenced by the local soil Se environment. Se is widely used as a supplement to livestock feeds, usually as sodium selenite/selenate. This practice has reduced geographic variation in the Se contents of animal food products. It does not, however, markedly increase tissue Se contents, as the incorporation of inorganic Se is limited to the synthesis of selenocysteine in the selenoproteins. Because both selenocysteine and selenomethionine are found mainly as constituents of proteins, the Se contents of foods of animal origin tend to be correlated with their protein content. Therefore, muscle meats from most species tend to contain 0.3–0.5mg Se/kg higher (4–15 times) amounts.

Human Diets. In most human diets, the dominant food sources of Se are cereals, meats and fish. In American diets fewer than two dozen foods provide most of the total dietary Se, with five foods contributing half: beef, bread, pork, chicken and eggs (Schubert et al., 1987). Dairy products contribute small amounts of Se to the total intakes in most countries, except in cases where their consumption is relatively high or where the rest of the diet provides little Se, or both (e.g., New Zealand). Vegetables and fruits are uniformly low in Se and provide only small amounts in most diets.

Because Se in foods is almost exclusively associated with protein, individuals with chronically low protein intakes are at risk for Se deficiency. Accordingly, Se-responsive muscular pain and weakness was been described in patients receiving long-term total parenteral nutrition without Se-supplementation (van Rij et al., 1979; Johnson et al., 1981; Stanley et al., 1982; Vinton et al., 1987; Terada et al., 1996), and protein-malnourished children with kwashiorkor or marasmus were found to have low plasma Se levels (Burk et al., 1967; Squali Houssaini et al., 1997; Ashour et al., 1999).

Selenium Intakes. Data from 68 countries indicated that suboptimal Se status (plasma/serum Se < 70–80ng/mL) is prevalent in almost half of those countries (Combs et al., 2001). Estimates of the average Se intakes in 18 countries varied from as low as 3µg/d in a Keshan disease-endemic part of China to as high as 224µg/d in Canada and Venezuela. The estimated intakes of several countries (Greece, Libya, New Guinea, Sweden, Turkey and the United Kingdom) fall below the U.S./Canadian RDA (Navarro-Alarcon and Cabrera-Vique, 2008). These differences are reflected in the average plasma/serum Se levels of residents of various countries, which span an order of magnitude (Combs et al., 2001).

This page titled 29.5: Food Sources (25b.5) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.6: Enteric Absorption (25b.6)

In general, food Se is well utilized. For the Se-amino acids, this involves digestion of their respective proteins, and absorption of the digestive products by active transport. For inorganic forms of Se, which do not require digestion, this involves absorption by simple (selenite) or carrier-mediated (selenate) diffusion (Wolffram et al., 1985; Turner et al., 1990; Van Dael et al., 2002). Wastney et al (2011) found that the enteric absorption of single doses (200µg) by humans was 98% for selenomethionine-Se and 84% for selenite-Se. Similar results were obtained by Thomson and Robinson (1986).

This page titled 29.6: Enteric Absorption (25b.6) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.7: Metabolism (25b.7)

Figure 25b.1. illustrates the metabolism of the major dietary forms of Se (selenomethionine, selenocysteine and, if supplemented, selenite). Selenium from each of these forms can be incorporated into functional selenoproteins as selenocysteine or non-specifically into general proteins as selenomethionine, and ultimately excreted as several methylated metabolites. selenocysteine and selenite can each be incorporated only into the selenoproteins by processes first involving reduction to selenide (Suzuki et al., 2007) followed by the *de novo*incorporation into selenocysteine during its biosynthesis via the co-translational conversion of tRNA-bound serine to selenocysteine. selenomethionine is incorporated differently; because it can charge the tRNAMet, seleno-methionine can be incorporated directly but nonspecifically into proteins in lieu of Met. For this reason, albumin is a primary component of plasma Se in Se-fed individuals. SeO2+
3(selenite and selenous acid (H_2SeO_3), and SeO2+
4 (selenite and selenic acid [H_2SeO_4]).

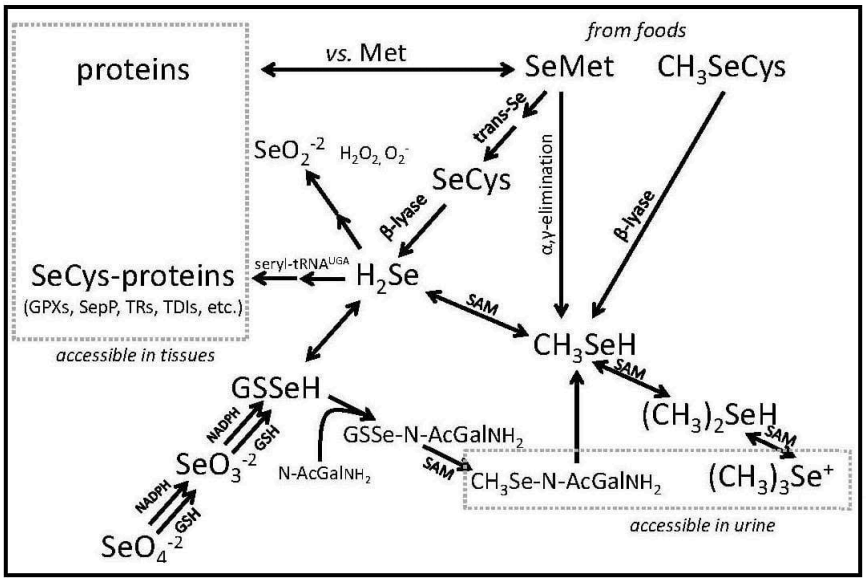


Figure 25b.1. Metabolism of Se. Note that inorganic Se can be incorporated into SeCys-containing selenoproteins by way of the obligate intermediate selenide (H_2Se); while Se from SeMet can also be incorporated into the SeCys-containing proteins by way of catabolism to H_2Se, and into other proteins directly by competing with methionine (Met) in protein synthesis. Other symbols:
SeO-2
2 = selenium dioxide; SeO-2
3 = selenite; SeO-2
4 = selenate; CH_3SeH = methyselenol; $(CH_3)_2SeH$ = dimethylselenide; $(CH_3)_3Se^+$ = trimethylselenonium ion Modified from Combs (2015), Nutrients, 7, 2209-2236,

Se is excreted by conversion to methylated metabolites Combs, 2016). The primary route of Se excretion in individuals with low-to-moderate Se intakes is across the kidney. The dominant urinary metabolite is a selenosugar, 1β-methylseleno-N-acetyl-galactosamine (Kobayashi et al., 2002). Individuals with relatively high Se intakes can also excrete trimethylselenonium or other selenosugars in the urine, as well as methylselenol in the breath. The latter is responsible for the “garlic breath” associated with selenosis in livestock. Metabolic tracer studies have shown that, over a 12-day period, Se-adequate adults excrete a total of 17% of Se from an oral selenite dose, but only 11% of Se from an oral selenomethionine dose, the longer turnover being due to the nonspecific incorporation of selenomethionine into body proteins (Patterson et al., 1989). Under normal circumstances, there appears to be only a small enterohepatic circulation of absorbed Se; therefore, fecal Se constitutes mostly unabsorbed dietary Se.

This page titled 29.7: Metabolism (25b.7) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.8: Biochemical Functions (25b.8)

The nutritional activities of Se are discharged by a group of selenoproteins, i.e., proteins containing selenocysteine. Selenoproteins are relatively rare in nature, occuring in bacteria, archaea, and eukaryota in varying numbers. None have been identified in fungi and plants. In eukaryotes, the greatest number of selenoproteins occurs in aquatic organisms (many fishes and algae have > 30 selenoproteins); 25 selenoproteins have been identified in humans (Gladyshev, 2016). Most, if not all, are involved in the regulation of redox signalling.

The best understood and most abundant members of the human selenoproteome include:

- Glutathione Peroxidases (GPXs), which reduce hydroperoxides using reducing equivalents from reduced glutathione (Flohé L and Brigelius-Flohé R, 2006).
- Selenoprotein P (SEPP), produced and excreted by the liver, functioning as the primary transporter of Se to peripheral tissues, and comprising 40–60% of total plasma Se (Burk and Hill, 2009).
- Iodothyronine Deiodinases (DIOs), which remove iodine from the thyroid hormones (T4, T3) in the metabolism of the active hormone, T3 (Köhrle, 2013).
- Thioredoxin Reductases (TRRs), NADPH-dependent flavoenzymes that function in intracellular redox regulation by reducing thioredoxin (Tamura and Stadtman, 1996).

Genomic searching (Kryukov et al., 2003; Shchedrina et al., 2010) has revealed additional selenoproteins:

- SepH: a nuclear protein of unknown function.
- SepI: a protein homologous to human choline/ethanolamine phosphotransferase with seven putative transmembrane domains.
- SepK, SePS : proteins of unknown function localized to endoplasmic reticulum (ER) and plasma membranes of mammalian cells.
- SepM, SepN, SepO, SepT: proteins of unknown function localized in the endoplasmic reticulum .
- SepV: a paralog of SepW apparently expressed only in the seminiferous tubules of the testes.

This page titled 29.8: Biochemical Functions (25b.8) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.9: Assessing Selenium Status (25b.9)

Se status refers to the amount of metabolically active or potentially metabolically active Se in the body. It includes the Se pool that is metabolically functional and, thus, the most nutritionally relevant, as well as other Se pools that can be mobilized to functional forms. It is a product of an individual's intake, retention and metabolism of Se-containing compounds. Accordingly, Se status has four components: Se intake, tissue Se, Se excretion and Se function (Figure 25b.2). (Combs, 2015).

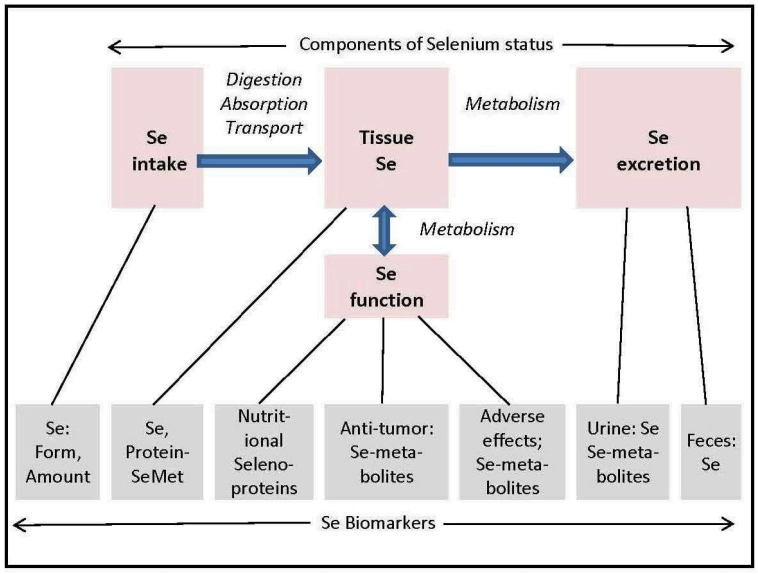


Figure 25b.2 Selenium biomarkers and their relationships to the components of selenium status . Modified from Combs (2015), Nutrients, 7, 2209–2236.

25b.9.1 Biomarkers of Selenium Intake

Se intakes of individuals or populations can be assessed as the sum of the products of the Se contents of particular foods and the amounts of those foods consumed. In practice, this is done using standard food composition tables and food frequency questionnaires, which may not yield accurate estimates of actual Se intake because of the geographic variation in food Se contents not captured in those tables. Therefore, determination of food Se by actual analysis of representative samples is necessary to produce valid Se intake estimates. Speciation of Se has been accomplished only for a few foods, e.g., lentil (Thavarajah et al., 2007).However, such information would not add value in understanding Se status; estimates of total Se intake are sufficient to determine whether individuals and populations meet the recommended Se intakes.

Imputing Se intake from measured Se biomarkers is most straightforward in subjects of deficient-to-low Se status whose plasma Se concentrations respond to supplemental Se in proportion to the magnitude of supplementation (Xia et al., 2005). For non-deficient subjects, however, the relationship of Se intake and plasma Se level depends on the form of Se consumed. Inorganic Se typically produces only minimal increases (< 20%) in subjects with plasma Se concentrations greater than about 70ng/mL (Duffield et al., 1999; Broome et al., 2004; Burk et al., 2006), while selenomethionine or selenomethionine-containing foods produce increases in plasma Se over a wide range, including subjects of relatively high Se status (Thomson and Robinson,1986; Duffield et al., 1999; Broome et al., 2004; Xia et al., 2005; Burk et al., 2006; Thavarajah et al., 2007; Ravn-Haren et al., 2008). The former effect reflects the fact that the central organ of Se metabolism, the liver, can convert inorganic Se to forms incorporated only into selenoproteins (Fig. 1), including two in the plasma, the extracellular Glutathione Peroxidase 3 and the Se-transporter Selenoprotein P that show plateaued expression at Se intakes recommended as being nutritionally adequate. The latter, continuing effect is due to the unregulated, non-specific incorporation of selenomethionine into albumin and other proteins over a virtually unlimited range of intakes (Thomson et al., 1982; 1986; Levander et al., 1983; Alfthan et al., 2000; Combs et al., 2012). For Se-nondeficient individuals consuming Se mostly as selenomethionine, Se intake can be predicted from the Se concentration of plasma.

25b.9.2 Biomarkers of Selenium Retention

Most absorbed Se is taken up by liver to re-enter the circulation as a component of Selenoprotein P (Burk and Hill, 2009). A significant amount(15–20%) is released in the urine over several days. Under normal conditions, there appears to be only a small enterohepatic circulation of absorbed Se; therefore, fecal Se is comprised mostly of the unabsorbed portion of ingested Se. In animals fed nutritionally adequate levels of selenite, fecal Se comprises about a fifth of Se excretion (Pedrosa et al., 2012); fecal Se increases at higher Se intakes.

Inferring Retention. Retention of Se in the body can be determined by the difference between the amount of Se ingested and the sum of urinary and fecal Se losses. For humans, this requires total collections of urine and stools, ideally over a few days. Alternatively, convenient samples of urine can be used if indexed to creatinine concentration to reduce errors associated with variation in urine output; a non-absorbable dietary marker that can be measured in the stool samples, can be used to calculate the amount of diet corresponding to the fecal sample analyzed for Se.

Assessment of Se retention is useful in livestock nutrition where it can be used as a measure of the bioavailability of feed forms of Se, or as a proxy for the amount of Se in edible tissues (Combs and Combs, 1986) Tissues retain Se largely in protein-bound forms, selenocysteine being specifically incorporated into selenoproteins and selenomethionine being nonspecifically incorporated into all proteins (Fig 1). Accordingly, animals fed dietary sources of selenomethionine will accumulate relatively large amounts of Se in both the specific and non-specific pools; whereas, those fed inorganic Se-compounds, e.g., selenite, will accumulate lesser amounts due to that form only entering the specific (selenocysteine) pool.

Estimating retained Se is useful in human studies as indicators of the degree to which an individual has met his/her Se requirement and may have Se reserves. This can be inferred from analyses of the two specific selenoproteins (Selenoprotein P, Glutathione Peroxidase 3) as well as the non-specific component of plasma Se. Those selenoproteins show maximal expression at plasma Se concentrations of 50–70ng/mL (Clausen and Nielsen, 1988).

Urinary Selenium. Excess absorbed Se can be eliminated in the urine by way of metabolism to methylated products including trimethylselenonium ion ($[CH_3]_3Se^+$) and a selenosugar 1β-methylseleno-N-acetyl-D-galactosamine. The dominant chemical form of urinary Se varies among species. In rodents, $TMSe^+$ can comprise 90% of urinary Se; in humans, the predominant form is the selenosugar (Ogra et al., 2002; Kobayashi et al., 2002; (Gammelgaard et al., 2003). In both rodents and humans, urinary Se excretion increases with increasing Se intake(Thomson et al., 1982; Combs et al., 2012; Pedrosa et al., 2012). Urinary Se excretion is related to both Se status, the form(s) of Se consumed(Robinson et al., 1997), and the availability of the methyl donor S-adenosylmethionine (Jackson et al., 2013). Urinary Se excretion is negatively associated with plasma homocysteine concentrations, but positively associated with plasma concentrations of folate and vitamin B_{12} (Combs et al., 2012). Urinary Se excretion of men and women differ, probably relating to sexually dimorphic selenoprotein biosynthesis as has been described in rodents (Stoedter et al., 2010). In a cohort of healthy, non-deficient Americans, the dose-dependent urinary Se excretion of women was 74% greater than that of men, despite the fact that both showed similar plasma Se responses to comparable intakes of Se (Thomson et al., 1982). This indicates that women do not retain selenomethionine-Se as well as men.

Urinary Se excretion also has a genetic determinant. The increase in urinary Se of non-deficient humans with increasing Se intake was 59% greater for individuals with the GPX1 679 T/T genotype compared to those with the GPX1 679 C/C genotype (Combs et al., 2011); while at baseline, the former individuals had slightly lower plasma Se levels. Because GPX1 accounts for some 60% of tissue Se (Combs et al., 2000),differences in its turnover are likely to affect Se retention and be manifest as differences in urinary Se excretion.

25b.9.3 Biomarkers of Tissue Selenium

Selenium can be retained in many organs. Greatest concentrations are found in the kidney, liver and pancreas, followed by cardiac and skeletal muscle; this pattern is remarkably similar across species, particularly under non-deficient conditions (Combs and Combs, 1986). Because of its non-specific incorporations into proteins, food sources of selenomethionine support greater tissue Se accumulation than comparable intakes of Se in inorganic forms. For this reason, animal studies using tissue Se accumulation as the measure of bioavailability typically show Se in foods/feedstuffs to be more highly available than selenite or selenate (Combs and Combs, 1986) .

Whole Blood. Earlier studies employed whole blood for the assessment of Se status. Those showed good correlations with functional biomarkers in cohorts of relatively low Se status, as well as responsiveness to supplementation with food sources of Se (largely selenomethionine). Blood samples can be dried and later analyzed by X-ray fluorescence analysis (Czuban et al., 2014).

Nevertheless, whole blood Se measures can be difficult to interpret in as much as they comprise both cellular and non-cellular constituents both of which have specific and non-specific components.

Plasma/Serum. The most useful tissue for assessing Se status, particularly in humans, is plasma (or serum, which is virtually equivalent with respect to Se-containing components and, therefore, total Se contents). Se is stable in plasma as long as microbial growth is prevented; it can be determined with good sensitivity and precision by automated atomic absorption spectrophotometry reducing interferences by using either graphite-furnace-AAS, or tandem inductively coupled plasma/mass spectrometry (ICP/MS). Plasma Se consists mostly of Se that has been incorporated both specifically as selenocysteine and non-specifically as selenomethionine into various proteins. Selenocysteine occurs in only two plasma selenoproteins, Selenoprotein P and Glutathione Peroxidase 3, which contain 10 and 4 selenocysteine residues, respectively. The synthesis of each, occurring in the liver and kidney, respectively, depends on an ample metabolic supply of selenide for the co-translational biosynthesis of selenocysteine (Carlson et al, 2006).

Plasma also contains a non-specific Se pool comprised of selenomethionine incorporated into proteins *in lieu* of Met with which it competes for $tRNA^{Met}$ binding. The potential for the unregulated, non-specific incorporation into proteins is a function of their various Met contents. In plasma, the most quantitatively important of these is albumin, which contains 6 Met residues per mole. Assuming that selenomethionine can compete for incorporation at each of those positions, selenomethionine protein would provide a plasma Met space equivalent to nearly 4mM. The non-specific pool can be estimated from the difference between the specific and total Se pools. By this approach, more than half of plasma Se of a non-deficient cohort was present in specific components (Glutathione Peroxidase 3 and Selenoprotein P comprising roughly equal amounts), with almost as much Se present in non-specific components (Combs et al., 2012). Further, the increase in plasma Se produced by selenomethionine-supplementation consisted entirely of expansion of the non-specific component, which reached 73% of plasma Se in subjects consuming some 300µg Se/day. Still, that amount comprised less than 0.05% of the potential selenomethionine space, indicating that the latter is unsaturable under practical conditions.

In contrast to selenocysteine and selenite, which enter the specific selenoprotein pool by a highly regulated process, the incorporation of selenomethionine into the nonspecific Se pool is unregulated, its size reflecting the intake of selenomethionine over a wide range. Accordingly, sources of selenomethionine are also more effective than inorganic Se in increasing tissue Se levels (Thomson et al., 1982; Levander et al., 1983; Thomson et al., 1993; Clausen and Nielsen, 1998; Duffield et al., 1999; Alfthan et al., 2000).

While Se intake is its primary determinant, plasma Se level can be also affected by other factors affecting any of the Se-containing proteins, e.g., gender, age, smoking status (Lloyd et al., 1983), protein malnutrition (Mathias and Jackson, 1982), pregnancy and lactation (Behne and Wolters, 1979; Levander et al., 1987), and inflammation (Maehora et al., 2002; Huang et al., 2012). Plasma Selenoprotein P levels have been found to be lower in obese individuals (BMI >30) (Combs et al., 2012). and in African Americans compared to white Americans (Hargreaves et al., 2014).

The informative value of plasma Se as a biomarker, therefore, depends on Se status (Figure 25b.3).

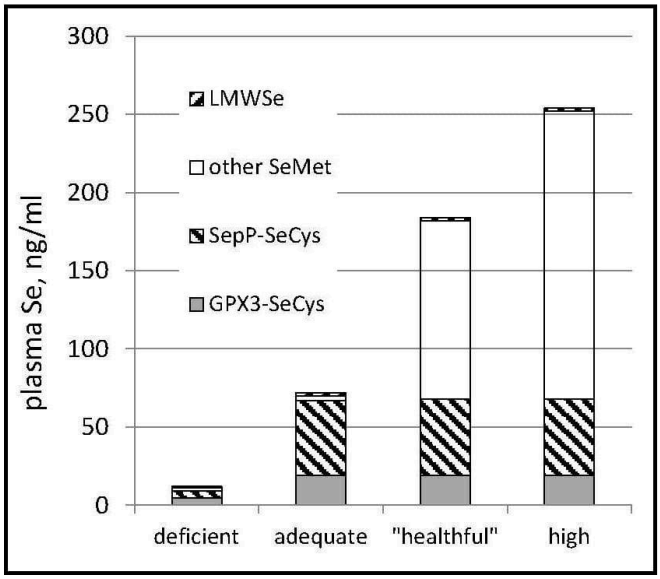


Figure 25b3. Stylized patterns of specific (solid areas, SeCys) and non-specific (hashed areas, presumed to be mostly SeMet in albumin) major components of plasma Se in subjects varying in Se status by consuming SeMet-containing foods. Not shown: Se occurring as SeMet in GPX3 and SEPP1, as their potential contributions to non-specific plasma Se are very small (< 0.1%). Group labels: deficient = less than maximal selenoprotein expression; adequate = maximal selenoprotein expression, i.e., meets nutritional requirement; "healthful" = supranutritional, including anti-tumorigenic doses; high = greater than needed for anti-tumorigenesis, but without adverse effects. Modified from Combs (2015), Nutrients, 7,2209–2236.

Measurements of Glutathione Peroxidase 3 activity and Selenoprotein P can identify the risk of nutritional deficiency. Measurements of total Se and, perhaps, the non-specific pool can identify luxus Se status. Thus, it is apparent that the largest contributor to the global variation in plasma Se, particularly among subjects with plasma Se concentrations greater than about 70ng/mL, is the non-specific pool.

For Se-nondeficient individuals consuming Se mostly as selenomethionine, Se intake can be predicted from the Se concentration of plasma. Such a prediction was made for a cohort of Americans with relatively high baseline plasma Se levels (142ng/mL) (Combs et al., 2012). using the amounts of Se administered as daily supplements of selenomethionine plus Se coming from the subjects' diets as estimated from a self-administered food frequency questionnaire and the USDA National Nutrient Database for Standard Reference, Release 27 (USDA, 2014) Results showed estimated Se intake expressed as μg/kg$^{0.75}$/d to be significantly correlated with measured Se-plasma, expressed as ng/mL. From that regression it is possible to predict Se intake:

$$\text{Se intake}_{\ \mu g/kg^{0.75}}/d = 0.44 + 0.03 \times \ \text{Se plasma}_{\ ng/mL}$$

Because virtually all circulating Se is bound to proteins, protein-deficient subjects will have low plasma/serum levels. This has been observed in children with kwashiorkor or marasmus (Mathias and Jackson, 1982), who may also have increased needs for Se and other antioxidant nutrients, due to the pro-oxidative effects of malnutrition and inflammation. Neonates typically have lower blood Se levels than their mothers(Lombeck et al., 1977; McKenzie et al., 1978; van Caillie-Bertrand et al., 1986), and low plasma Se levels have been associated with increased risk of respiratory morbidity among low-birth-weight newborns. Low blood Se levels have been observed in infants with phenylketonuria, but these effects have been due to the use of parenteral feeding solutions that had not been supplemented with Se.

Cells. Three cell types are accessible for use in assessing tissue Se status in humans: erythrocytes, lymphocytes and buccal cells. The Se contents of each respond to selenomethionine supplementation. Erythrocyte Se requires digestion after which it can be determined by GF-AAS. Because erythrocytes are collected with blood, they seldom offer advantages over plasma as a specimen for assessing Se status based on Se content. Buccal cells offer the possibility of a minimally invasive biomarker of somatic cell Se status. These cells can easily be collected in amounts sufficient for analyzing Se and can be acid-solubilized rather than digested prior to GF-AAS analysis. Nevertheless, few groups to have used buccal cells for this purpose (Thomson et al., 1982; O'Callaghan et al., 2008; Combs et al., 2012).. In a non-deficient cohort, buccal Se concentration was positively associated with age and nutritional supplement use, and to respond to selenomethionine supplementation.

Hair and Nail Selenium. Because hair and nails contain sulfur-rich proteins, each can accumulate Se. In addition, hair and nail specimens are stable and can be collected non-invasively, and can be stored at very low cost. For these reasons, these specimens have been used to assess Se status in epidemiological studies. Such studies have shown the amounts of Se in hair to be greater in individuals with high Se intakes (Steven Morris et al., 1983), and increase with Se-supplementation (Hunter et al., 1990). Very high hair and nail Se levels were observed among residents of seleniferous locales (Yang et al., 1983), in cases of Se-poisoning (Lombeck et al., 1987), and in users of Se-containing antidandruff shampoos (Davies, 1982; LeBlanc et al., 1999).

The analysis of Se in hair and nails calls for careful cleaning, with special concern for hair as to whether subjects may have used Se-containing, anti-dandruff shampoos. Selenium can be determined in these matrices by neutron activation analysis, and electrothermal (graphite furnace) atomic absorption spectrophotometry (GF-AAS) after acid-digestion. With standardized procedures, e.g., Bank et al., (1981) for collecting these samples, both hair and nail Se levels correlate well with blood/plasma Se concentration (Yang et al., 1989b, 1989c; Swanson et al., 1990; Longnecker et al., 1996).

It bears pointing out that the use of these specimens to assess Se status relies on the implicit but never validated assumption that their Se contents reflect some metabolically relevant component of body Se. In fact, hair/nail Se is not in dynamic equilibrium with any component of circulating Se; it is actually an excretory form of the element. As such, samples of each contain Se secreted at some time past, presumably reflecting Se status at that time and making them most useful in studies of populations with stable dietary practices. These contributions are very small in comparison to urinary Se output and can be ignored in estimating Se retention.

25b.9.4 Biomarkers of Selenium Function

In principle, measurement of any of the selenoproteins would yield information immediately relevant to the physiological function of Se as an essential nutrient. The most widely used as biomarkers in animal studies are the GPXs. These, the thioredoxin reductasess, and the iodothyronine deiodinasess show maximal expression in most species with dietary Se concentrations of 0.1–0.2mg/kg (Yang et al., 1989; Sunde, 1990; Sunde and Raines, 2011). Only one selenoprotein, SepW, is known not to be maximally expressed under such conditions, but to be upregulated by doses of Se in the anti-tumorigenic range (Hawkes et al., 2009).

Selenoenzymes. The list of practical biomarkers of Se function is much shorter, being limited to those occuring in accessible tissues: The extracellular isoform of GPX (Glutathione Peroxidase 3), which comprises 10–25% of plasma Se (lower levels being associated with greater plasma Se concentrations); Selenoprotein P, which comprises 20–70% of plasma Se (lower percentages being associated with greater plasma Se concentrations); and the intracellular isoform (GPX1), which can be assayed in erythrocytes, lymphocytes, buccal cells and tissue biopsy specimens.

Activities of the GPXs can be measured spectrophotometrically using a glutathione reductase-coupled assay (Lawrence and Burk, 1976). Because heme iron can cause interfering glutathione oxidation in that assay, it is necessary to treat erythrocyte samples with Drabkin's reagent when assessing GPX1. The GPXs are homotetramers; each 23kD subunit contains a selenocysteine residue at its active center. Unlike GPX1, Glutathione Peroxidase 3 is a glycoprotein produced mostly by the kidney. Selenoprotein P is present in plasma as multiple variants with an average molecular weight of 60kD (Broderick et al., 1987; Steinbrenner et al., 2007; Méplan et al., 2009; Chittum et al., 1996). Although its genetic coding would suggest as many as 10 selenocysteine residues per mole (Hill et al., 1993), studies in the rat have shown only 7–8 residuesper mole (Saito et a, 2014). Selenoprotein P can be measured by immunoassay (Hollenbach et al., 2008; Tanaka et al., 2016).

The dependency of the expression of these selenoproteins on Se intake has been demonstrated in humans for Glutathione Peroxidase 3 (Yang et al., 1987; Xia, et al., 1989; Xia et al., 2005) and Selenoprotein P (Hill et al., 1996; Xia et al., 2005; Xia et al., 2010) in studies conducted in parts of China with endemic Se deficiency. In both cases, pooled analyses of subjects from different communities with differing plasma Se concentrations showed Glutathione Peroxidase 3 activity and plasma Selenoprotein P1 concentration each to be directly related to plasma Se concentration over ranges of about 8–80ng Se/mL. This work has provided the basis for setting dietary Se requirements, which have been based on either two-thirds or maximal Glutathione Peroxidase 3 expression, yielding dietary recommendations for adults from 26–70µg Se/d (Hurst et al., 2013).

Maximal Glutathione Peroxidase 3 expression may occur at a lower level of Se status than maximal Selenoprotein P expression. Xia et al (2020) found that Se-deficient subjects supplemented with selenomethionine for 20wks reached maximal Glutathione Peroxidase 3 activities with plasma Se levels about 66ng/mL, but that their plasma Selenoprotein P concentrations were not maximized until their plasma Se levels were at least a third higher. However, Selenoprotein P may have taken longer to reach new steady-state levels, as after 40wks of intervention the threshold of maximal expression of Selenoprotein P approached that of Glutathione Peroxidase 3 both being in the range of 35–45µg Se/d corresponding to plasma Se concentrations of 50–70ng/

mL (Nève, 1995). These thresholds are typically exceeded in western populations; Nève (1995). noted that subjects with plasma Se concentations > 70ng/mL do not show Glutathione Peroxidase 3 responses to Se-supplementation. Accordingly, negative Glutathione Peroxidase 3 and Selenoprotein P responses to Se supplementation are typical in trials conducted in cohorts with greater Se intakes such as in the US (Burk et al., 2006; Combs et al., 2012) and UK (Hurst, et al., 2010).

Obesity can affect Selenoprotein P expression. In a cohort of non-deficient subjects, plasma Selenoprotein P levels were slightly depressed among subjects with body mass index (BMI) greater than 30. Méplan et al (2007) also found relatively low-Se (mean plasma Se 90ng/mL) subjects with BMI> 30 to have lower baseline plasma Se concentrations than non-obese subjects; obese females also showed lower Glutathione Peroxidase 3 activities. These findings may reflect influences of pro-inflammatory cytokines, which have been found to inhibit Selenoprotein P1 expression (Dreher et al., 1997), or dysregulation of gluconeogenesis, as Selenoprotein P appears to be regulated as a gluconeogenic enzyme(Speckmann et al., 2008; 2009).

Genetic Variability and gender contribute to variance in biomarkers of Se function. Individuals with GPX1 679T/T genotype (which has been associated with increased cancer risks) (Ratnasinghe et al., 2000; Hu and Diamond, 2003) had lower plasma Se levels than those with GPX1 679C/C alleles; those with the *Selenoprotein P1* 24731 A/A genotype had 27% greater plasma Selenoprotein P levels than those with the *Selenoprotein P1* 24731 G/A or G/G genotypes; and those with the SEP15 811 T/C genotype had greater buccal cell Se levels than those with the SEP15 811 C/C genotype (Combs et al., 2011). Hurst et al (2010). found two variants of Selenoprotein P to affect circulating concentrations of selenoproteins. One variant at position 234 of the protein may affect Selenoprotein P turnover; the other variant at r25191 in the 3'-untranslated region may affect Se incorporation into Selenoprotein P. They found women with the *Selenoprotein P1* 234GA genotype showed lower Selenoprotein P levels than men of the same genotype; while women with the *Selenoprotein P1* 234GG genotype showed lower Selenoprotein; P levels than men of that genotype. Subjects with the *Selenoprotein P1* 234AA genotype showed lower plasma TR1 concentrations compared to those of the other genotypes. Variants in r25191 did not affect *Selenoprotein P1* levels; however, individuals with the r25191GG genotype showed greater plasma TR1 concentrations than women of that genotype, while males of the r25191GA genotype were the only subjects to show TR1 responses to Se supplementation.

Non-specific parameters. Several non-specific parameters have been used to assess metabolic impacts of Se deficiency. The plasma triiodothyronine / thyroxine (T3:T4) ratio ratio, normally about 20:1, is elevated by Se-deficiency (Vanderpas et al., 1990; 1993) reflecting limited expression of the iodothyronine deiodinasess. While useful as an indicator of thyroid hormone status, the T3:T4 ratio lacks specificity for Se status, as not only can the conversion of thyroxine to triiodothyronine be impaired by Se deprivation, but also the synthesis of thyroxine can be impaired by deprivation of Se and/or iodine. That the pattern of low molecular weight Se-metabolites in the hindgut lumen corresponds with that of feces, would appear to suggest that a fecal sample may ve useful as an indicator of the Se in the colon.

25b.9.5 Biomarkers Relevant to Anti-tumorigenesis.

Studies with animal models have demonstrated anti-tumorigenic effects of Se involving several aspects of cell proliferation, migration and apoptosis (Combs and Yan, 2016). These effects are manifest with dietary Se levels of at least 1.5mg/kg, while most selenoproteins are maximally expressed at far lower dietary levels (0.1–0.2mg/kg). For that reason, the anti-tumorigenic effects of supranutritional doses of Se would appear to involve functions of Se-metabolites and not of selenoproteins (Combs and Gray, 1998). Still, that certain selenoproteins may play roles in Se-antitumorigenesis is suggested by the finding of Hawkes et al. (2009) that the mRNA for SEPW1 is up-regulated by supranutritional Se treatment, and that silencing SEPW1 is associated with reduced stability of the tumor suppressor protein p53, suggesting that SEPW1 may serve in its stabilization (Hawkes et al., 2012).

Evidence from human trials is consistent with the view of anti-tumorigenic Se-metabolites, such as methylselenol (CH_3SeH), showing reduction in cancer risk among subjects of nutritionally adequate (i.e., maximal Glutathione Peroxidase 3 expression), but not high Se status. For example, the Nutritional Prevention of Cancer Trial (Combs et al., 2001; Duffield-Lillico et al., 2003) found supplemental Se to reduce cancer risk in non-deficient Americans only if their baseline plasma Se levels were less than about 120ng/mL. This is consistent with the results of the larger Selenium and Vitamin E Cancer Prevention Trial (SELECT) (Lippman et al., 2009), which found no risk reduction in a cohort with baseline plasma Se concentrations exceeding that level. Although CH3SeH and, perhaps, other methylated Se-metabolites would be candidate biomarkers relevant to Se-antitumorigenesis, none have been reported in biological specimens. At the moment, therefore, the most useful biomarker relevant to Se-antitumorigenesis is plasma Se, the target level being about 120ng/mL above which current evidence indicates no further risk reduction.

25b.9.6 Biomarkers Relevant to Adverse Effects.

Biomarkers Relevant to Adverse Effects. The potential for very high Se status to produce adverse physiological effects has been established from animal studies and accidental exposures of humans. These reports have produced an array of clinical indicators but few biomarkers with predictive value. For this reason, the default choice has been to use as risk indicators the highest Se tissue levels observed with no adverse effects. Most studies show no adverse effects in human subjects with plasma Se levels < 1000ng/mL.

This page titled 29.9: Assessing Selenium Status (25b.9) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

29.10: Questions Remain (25b.10)

Understanding of Se, the most recently recognized essential nutrient, continues to grow. It has become clear that consequences of Se deprivation are typically sub-clinical in nature, requiring other precipitating factors (deficiencies of vitamin E, iodine and/or protein; exposure to viruses, or carcinogens) to reveal physiological effects of sub-optimal expression of selenoenzymes. It is also apparent that supranutritional intakes of Se, at which selenoenzymes are fully expressed, can be beneficial. These facts raise three pressing questions:

- Can Se prevent cancer in humans? The clinical significance of Se in cancer prevention remains a subject of debate. The nine randomized clinical trials conducted to date have yielded inconsistent results. Systematic reviews have found that Se may be effective in preventing cancer in individuals of low to adequate, but not high, Se status; and a U-shaped dose-response relationship has been proposed (Chiang et al., 2009). These conclusions are weakened by the fact that several trials have not followed robust protocols, and none has been conducted in recent years or in cohorts with individuals of both marginal and adequate Se status.
- What is the window of safety for Se? While it is clear that very high levels of exposure are required to produce selenosis, resolving the question of whether supranutritional intakes of Se can increase T2D risk remains unclear. Clarifying that will call for well controlled trials with subjects randomized by T2D risk factors (high BMI, elevated fasting glucose) and followed with unequivocal diagnostic indicators (fasting glucose, HbA1c, oral glucose tolerance). Until then, there would appear to be no justification for any healthy adult, regardless of his/her baseline Se status, to consume more than 100µg Se/day.
- Who Can Benefit from Se? This is the ultimate question. It is clear that supplementation of adults with basal Se intakes < 50µg/d will increase selenoenzyme expression, and that the magnitude of those increases will depend on the extent to which baseline plasma Se levels were below about 90ng/mL. That individuals with plasma Se levels less than about 120ng/mL may also benefit from increased Se intakes is suggested from results of the NPC Trial in which supplemental Se reduced cancer risk among adults with plasma Se levels in the range of80–120ng/mL. This suggests that individuals with plasma Se levels in that range may benefit from cancer protection unrelated to selenoenzyme expression. If so, many adults may benefit from increased Se intakes: some to maximize selenoprotein expression, others to reduce cancer risks.

This page titled 29.10: Questions Remain (25b.10) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

INDEX

A

abdominal cramping
- 25.6: Effects of high magnesium intakes (23c.6)

absorption
- 14.1: Introduction and functions of iron (17.1)
- 17.3: Absorption and metabolism of vitamin C (19.3)
- 19.1: Riboflavin (20b.1)
- 21.1: Folate (22a.1)
- 22: Vitamin B12 (Chapter 22b)
- 22.1: Introduction (22b.1)
- 23.4: Dietary sources and intakes of calcium (23a.4)
- 25.3: Absorption and metabolism (23c.3)
- 26.3: Absorption and metabolism (23c.3)
- 28.1: Zinc (24c.1)
- 29.6: Enteric Absorption (25b.6)

abundance
- 23.1: Calcium (23a.1)

accuracy
- 1.6: Important characteristics of assessment measures
- 12.9: Bioelectrical impedance (14.8)
- 16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
- 4.5: Summary
- 8.3: Errors in anthropometry

accurate measurement
- 12.8: Total body electrical conductivity (14.7)

active transport
- 25.3: Absorption and metabolism (23c.3)

acute kidney injury
- 24.8: Abnormalities in serum phosphate (23b.8)

adequacy
- 6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

adequate intake
- 5.1: Development of Nutrient Reference Values (8a.1)
- 6: Evaluation of intakes and diets (Chapter 8b)

adipose tissue
- 12.10: Computerized tomography (14.9)
- 12.11: Magnetic resonance imaging (14.10)
- 12.13: Ultrasound (14.12)
- 13.3: Biomarkers of status (15.3)

adiposity
- 12.14: Summary - Method comparisons (14.13)

adolescents
- 9.5: BMI in children and adolescents (10.4)

adult
- 12.2: Chemical analysis of cadavers (14.1)

adult nutrition
- 26.5: Food sources and dietary Intakes (23c.5)

adults
- 23.5: Nutrient reference values for calcium (23a.5)
- 24.3: Interpretive criterion- serum phosphate (23b.3)
- 24.6: Nutrient reference values (23b.6)
- 25.5: Food sources and dietary Intakes
- 25.7: Occurrence of dietary deficiency (23c.7)
- 27.7: Diamine oxidase (24b.6)
- 28.13: Taste acuity (24c.13)
- 28.5: Urine zinc (24c.5)
- 29.4: Selenium Requirements (25b.4)
- 9.2: Measurements of body size (10.1)

aging
- 20.1: Niacin - Introduction (20c.1)

agricultural data
- 2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)

agricultural practices
- 29.5: Food Sources (25b.5)

agriculturaldata
- 2.2: Food Balance Sheets

agriculture
- 24.1: Phosphorus (23b.1)
- 7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)

AHEI
- 7.4: Alternate Healthy Eating Index (8c.4)

alcoholics
- 17.4: Deficiency of vitamin C in humans (19.4)

alkaline phosphatase
- 16.10: Other vitamin D metabolites (18b.11)
- 28.14: Zinc-dependent enzymes (24c.14)

allosteric activation
- 25.2: Functions of magnesium (23c.2)
- 26.2: Functions of magnesium (23c.2)

amino acid metabolism
- 21.3: Conclusions (22a.3)

amino acids
- 29.2: Selenium Chemistry (25b.2)
- 29.6: Enteric Absorption (25b.6)

analysis
- 17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
- 24.2: Biological forms of phosphate and their measurement (23b.2)

analytical methods
- 16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
- 28.6: Hair zinc concentrations (24c.6)

analytical techniques
- 28.7: Nail zinc (24c.7)

anemia
- 13.6: Evaluation of the selected nutritional biomarker (15.6)
- 14: Iron (Chapter 17)
- 14.10: Multiple biomarkers (17.10)
- 14.11: Emerging iron status indicators (17.11)
- 14.2: Hemoglobin (17.2)
- 14.3: Hematocrit or packed cell volume
- 14.4: Red-cell indices
- 14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
- 14.7: Serum ferritin (17.7)
- 14.8: Erythrocyte protoporphyrin
- 14.9: Serum soluble transferrin receptor (17.9)
- 19.2: Biomarkers of riboflavin status (20b.2)
- 21.2: Biomarkers of folate status (Chapter 22a.2)
- 22.2: Biomarkers of vitamin B12 status (22b.2)
- 27.3: Serum ceruloplasmin (24b.2)

animal foods
- 29.5: Food Sources (25b.5)

animal protein
- 24.7: Dietary sources

anisocytosis
- 14.5: Red-cell distribution width (17.5)

Anorexia Nervosa
- 12.14: Summary - Method comparisons (14.13)
- 12.6: In vivo activation analysis (14.5)

anthropometric
- 10: Body composition (Chapter 11)
- 11: Evaluation of Anthropometric Data (Chapter 13)
- 11.1: Introduction (13.0)
- 8.1: Measurements, indices, and indicators
- 9.1: Anthropometric assessment of body size (10.0)
- 9.3: Growth indices, indicators, and recommended growth reference data (10.2)

anthropometric indices
- 1.7: Evaluation of nutritional assessment indices

anthropometry
- 1.4: Nutritional assessment indices and indicators
- 10.1: Anthropometric assessment of body composition (11.0)
- 10.2: Assessment of body fat (11.1)
- 10.3: Assessment of fat-free mass (11.2)
- 11.2: Anthropometric indices and their modes of expression (13.1)
- 11.3: Use of anthropometric indices in clinical settings (13.2)
- 11.4: Screening systems to identify individuals at risk in a population (13.3)
- 11.5: Applications of anthropometry in public health (13.4)
- 28.4: Linear growth (24c.4)
- 8: Introduction to Anthropometry (Chapter 9)
- 8.2: Advantages and limitations of anthropometry
- 8.4: Interpretation and evaluation of anthropometric data

antioxidant
- 17: Vitamin C (Chapater 19)
- 17.2: Functions of Vitamin C (19.2)

antioxidant protection
- 1.1: New developments in nutritional assessment

antioxidant status
- 27.1: Copper (24b.0)

antioxidants
15.1: Introduction (18a.1)
17.8: Plasma ascorbic acid (19.8)
19: Riboflavin (Chapter 20b)
19.1: Riboflavin (20b.1)
28.11: Oxidative stress and DNA integrity (24c.11)

apoptosis
16.1: Functions of vitamin D (18b.1)

appetite
28: Zinc (Chapter 24c)

appropriate method
3: Food consumption of individuals

aquatic organisms
29.8: Biochemical Functions (25b.8)

archaea
29.8: Biochemical Functions (25b.8)

ascorbic acid
17.1: Vitamin C (19.1)
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
17.12: Ascorbic acid in other cells and fluids (19.12)
17.13: Body pool size (19.13)
17.8: Plasma ascorbic acid (19.8)

ascorbicAcid
17: Vitamin C (Chapater 19)
17.14: Functional tests of vitamin C status (19.14)
17.2: Functions of Vitamin C (19.2)
17.7: Indices of vitamin C status (19.7)

assessment
13: Biomarkers (Chapter 15)
19: Riboflavin (Chapter 20b)
2: Food available for consumption at National and Household Levels
25: Phosphorus (Chapter 23b)
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)
7: Dietary guidelines and quality (Chapter 8c)

ATP synthesis
25.2: Functions of magnesium (23c.2)
26.2: Functions of magnesium (23c.2)

atrial fibrillation
26.4: Magnesium deficiency in humans (23c.4)

ATX1
27.8: Copper chaperones (24b.7)

averageRequirements
17.6: Nutrient reference values for vitamin C (19.6)

B

B12 deficiency
22.2: Biomarkers of vitamin B12 status (22b.2)

B6
21.1: Folate (22a.1)

bacteria
29.8: Biochemical Functions (25b.8)

balance
2: Food available for consumption at National and Household Levels
24.4: Phosphate balance (23b.4)

Bariatric Surgery
12.1: An introduction to techniques used to measure body composition (14.0)

beriberi
18: Thiamine (Chapter 20a)
18.1: Thiamine (20a.1)
18.3: Conclusions (20a.3)

BIA
12.9: Bioelectrical impedance (14.8)

bias
4: Measurement Error in Dietary assessment (Chapter 5)
4.1: Measurement error in dietary intake data

binding ligands
28.15: Erythrocyte zinc (24c.15)

bioactive constituents
13.1: Biomarkers to assess nutritional status (15.1)

bioavailability
13.2: Biomarkers of exposure (15.2)
17.5: Food sources and dietary intakes of vitamin C (19.5)
20: Niacin (Chapter 20c)
21.1: Folate (22a.1)
22: Vitamin B12 (Chapter 22b)
23.4: Dietary sources and intakes of calcium (23a.4)
24.7: Dietary sources
28.2: Dietary Zinc Intake (24c.2)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

biochemical
16.12: Calcium and phosphorus in serum and urine (18b.14)

biochemical changes
28.14: Zinc-dependent enzymes (24c.14)

biochemical markers
16.10: Other vitamin D metabolites (18b.11)
27.4: Erythrocyte superoxide dismutase (24b.3)

biochemicalTests
17.7: Indices of vitamin C status (19.7)

biochemistry
29.2: Selenium Chemistry (25b.2)

biodiversity loss
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)

bioelectrical impedance
12.1: An introduction to techniques used to measure body composition (14.0)
12.8: Total body electrical conductivity (14.7)

bioelectrical impedance analysis
12: Body Composition- Laboratory Methods (Chapter 14)

bioelectricalimpedance
12.9: Bioelectrical impedance (14.8)

bioindicator
28.13: Taste acuity (24c.13)

bioinformatics
13.7: Application of new technologies (15.7)

biological activity
17.1: Vitamin C (19.1)

biological fluids
13.3: Biomarkers of status (15.3)
24.2: Biological forms of phosphate and their measurement (23b.2)

biological sample
12.5: Multiple dilution methods (14.4)

biomarker
16.5: Nutrient reference values for vitamin D (18b.5)
18.2: Biochemical indices of thiamine status (20a.2)
18.3: Conclusions (20a.3)
24.3: Interpretive criterion- serum phosphate (23b.3)
27.9: Other indices of copper status (24b.8)
28.11: Oxidative stress and DNA integrity (24c.11)
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
28.18: Multiple biomarkers (24c.18)
28.3: Serum zinc (24c.3)
28.5: Urine zinc (24c.5)
28.6: Hair zinc concentrations (24c.6)
28.7: Nail zinc (24c.7)
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)

biomarkers
1.1: New developments in nutritional assessment

biomarkers
1.7: Evaluation of nutritional assessment indices
13: Biomarkers (Chapter 15)
13.2: Biomarkers of exposure (15.2)
13.3: Biomarkers of status (15.3)
13.4: Biomarkers of function (15.4)
13.5: Factors affecting choice of nutritional biomarkers (15.5)
14: Iron (Chapter 17)
14.10: Multiple biomarkers (17.10)
14.11: Emerging iron status indicators (17.11)
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
14.7: Serum ferritin (17.7)
14.9: Serum soluble transferrin receptor (17.9)
15.2: Indices of vitamin A status (18a.2)
16.7: Interpretive criteria (18b.8)
19: Riboflavin (Chapter 20b)
19.2: Biomarkers of riboflavin status (20b.2)
20.2: Indices of niacin status (20c.2)
21.2: Biomarkers of folate status (Chapter 22a.2)
22.2: Biomarkers of vitamin B12 status (22b.2)
27.2: Serum copper (24b.1)
27.3: Serum ceruloplasmin (24b.2)
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
28.14: Zinc-dependent enzymes (24c.14)
28.16: Leukocyte zinc (24c.16)
28.17: Salivary zinc (24c.17)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)
29.9: Assessing Selenium Status (25b.9)
7.8: Dietary Inflammatory Indices (8c.8)

biosynthetic pathway
17.1: Vitamin C (19.1)

blood
13.3: Biomarkers of status (15.3)
17.13: Body pool size (19.13)
23: Calcium (Chapter 23a)

blood analysis
14.3: Hematocrit or packed cell volume

blood cell types
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)

blood cells
14.2: Hemoglobin (17.2)
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)

blood measurement
14.8: Erythrocyte protoporphyrin

blood pressure
7.6: Dietary Approaches to Stop Hypertension scores
blood thiamine diphosphate
18: Thiamine (Chapter 20a)
blood_cells
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
bloodThDP
18.3: Conclusions (20a.3)
BMD
23.9: Measurement of bone mineral content and bone mineral density (23a.9)
BMI
1.4: Nutritional assessment indices and indicators
11.1: Introduction (13.1)
11.3: Use of anthropometric indices in clinical settings (13.2)
11.4: Screening systems to identify individuals at risk in a population (13.3)
12.14: Summary - Method comparisons (14.13)
9: Body size (Chapter 10)
9.1: Anthropometric assessment of body size (10.1)
9.4: Body mass index in adults (10.3)
9.5: BMI in children and adolescents (10.4)
body
23.1: Calcium (23a.1)
body cell mass
12.3: Total body potassium (TBK) (14.2)
body composition
10.1: Anthropometric assessment of body composition (11.1)
10.2: Assessment of body fat (11.1)
10.3: Assessment of fat-free mass (11.2)
11: Evaluation of Anthropometric Data (Chapter 13)
11.2: Anthropometric indices and their modes of expression (13.1)
12: Body Composition- Laboratory Methods (Chapter 14)
12.1: An introduction to techniques used to measure body composition (14.1)
12.10: Computerized tomography (14.9)
12.11: Magnetic resonance imaging (14.10)
12.12: Dual energy X-ray absorptiometry (14.11)
12.13: Ultrasound (14.12)
12.14: Summary - Method comparisons (14.13)
12.2: Chemical analysis of cadavers (14.1)
12.3: Total body potassium (TBK) (14.2)
12.4: Total body water from isotope dilution (14.3)
12.6: In vivo activation analysis (14.5)
12.7: Densitometry (14.6)
12.8: Total body electrical conductivity (14.7)
8: Introduction to Anthropometry (Chapter 9)
8.1: Measurements, indices, and indicators
8.2: Advantages and limitations of anthropometry
8.3: Errors in anthropometry
9.4: Body mass index in adults (10.3)
body fat
10.2: Assessment of body fat (11.1)
12.1: An introduction to techniques used to measure body composition (14.1)
body fluid compartments
12.5: Multiple dilution methods (14.4)
body fluids
17.12: Ascorbic acid in other cells and fluids (19.12)
23.1: Calcium (23a.1)
body mass index
12.7: Densitometry (14.6)
8.1: Measurements, indices, and indicators
body measurements
8: Introduction to Anthropometry (Chapter 9)
body pool size
17.13: Body pool size (19.13)
body pools
25.1: Introduction (23c.1)
26.1: Introduction (23c.1)
body size
8.1: Measurements, indices, and indicators
9: Body size (Chapter 10)
9.1: Anthropometric assessment of body size (10.1)
9.2: Measurements of body size (10.1)
bodycomposition
10: Body composition (Chapter 11)
12.9: Bioelectrical impedance (14.8)
bodyfat
10: Body composition (Chapter 11)
12.9: Bioelectrical impedance (14.8)
bone
23.1: Calcium (23a.1)
24.5: Hormonal regulation of serum phosphate (23b.5)
bone health
16: Vitamin D (Chapter 18b)
16.10: Other vitamin D metabolites (18b.11)
23.6: Calcium deficiency (23a.6)
bone measurement
12.13: Ultrasound (14.12)
bone mineral
26.1: Introduction (23c.1)
bone mineral content
23.9: Measurement of bone mineral content and bone mineral density (23a.9)
bone remodeling
23.2: Functions of calcium (23a.2)
23.3: Absorption and metabolism of calcium (23a.3)
23.8: Biochemical indices of calcium status (23a.8)
bonehealth
16.5: Nutrient reference values for vitamin D (18b.5)
17.4: Deficiency of vitamin C in humans (19.4)
boneMetabolism
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)
boneMineralization
16.1: Functions of vitamin D (18b.1)
bones
16.3: Vitamin D deficiency in humans (18b.3)
23: Calcium (Chapter 23a)
23.2: Functions of calcium (23a.2)
brain size
9.2: Measurements of body size (10.1)

C

cadaverine
27.7: Diamine oxidase (24b.6)
cadavers
12.2: Chemical analysis of cadavers (14.1)
calciferol
16: Vitamin D (Chapter 18b)
calcitriol
16: Vitamin D (Chapter 18b)
16.1: Functions of vitamin D (18b.1)
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)
23: Calcium (Chapter 23a)
24.5: Hormonal regulation of serum phosphate (23b.5)
calcium
16.1: Functions of vitamin D (18b.1)
16.12: Calcium and phosphorus in serum and urine (18b.14)
16.4: Food sources and dietary intakes (18b.4)
16.6: Biomarkers of vitamin D status (18b.6)
23: Calcium (Chapter 23a)
23.1: Calcium (23a.1)
23.2: Functions of calcium (23a.2)
23.3: Absorption and metabolism of calcium (23a.3)
23.4: Dietary sources and intakes of calcium (23a.4)
23.5: Nutrient reference values for calcium (23a.5)
23.6: Calcium deficiency (23a.6)
23.7: Effects of high intakes of calcium (23a.7)
23.8: Biochemical indices of calcium status (23a.8)
25.3: Absorption and metabolism (23c.3)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)
calcium balance
23.9: Measurement of bone mineral content and bone mineral density (23a.9)
calcium intake
23.6: Calcium deficiency (23a.6)
calcium malabsorption
23.3: Absorption and metabolism of calcium (23a.3)
calcium metabolism
23.8: Biochemical indices of calcium status (23a.8)
calcium phosphate
24.2: Biological forms of phosphate and their measurement (23b.2)
calcium_absorption
16.11: Serum parathyroid hormone (18b.13)
calciumMetabolism
16.2: Metabolism of vitamin D (18b.2)
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)
calibration
12.3: Total body potassium (TBK) (14.2)
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
caloric intake
7.9: Diet quality based on level of food processing (8c.9)
cancer
12.3: Total body potassium (TBK) (14.2)
29.1: Overview (25b.1)
cancer prevention
29: Selenium (Chapter 25b)
29.10: Questions Remain (25b.10)
cancer risk
29: Selenium (Chapter 25b)
29.3: Health Effects (25b.3)
29.9: Assessing Selenium Status (25b.9)
carbohydrates
5.6: European Dietary Reference Values (8a.6)

carbonic anhydrase
28.15: Erythrocyte zinc (24c.15)
cardiac abnormalities
27.6: Skin lysyl oxidase (24b.5)
cardiovascular
17.2: Functions of Vitamin C (19.2)
25.4: Magnesium deficiency in humans
cardiovascular disease
10.2: Assessment of body fat (11.1)
12.14: Summary - Method comparisons (14.13)
21.2: Biomarkers of folate status (Chapter 22a.2)
23.6: Calcium deficiency (23a.6)
23.7: Effects of high intakes of calcium (23a.7)
24.7: Dietary sources
26.4: Magnesium deficiency in humans (23c.4)
26.8: Indices of magnesium status (23c.8)
7.4: Alternate Healthy Eating Index (8c.4)
7.6: Dietary Approaches to Stop Hypertension scores
9.4: Body mass index in adults (10.3)
carotenoids
15.1: Introduction (18a.1)
catabolism
17.3: Absorption and metabolism of vitamin C (19.3)
cation
23: Calcium (Chapter 23a)
25.2: Functions of magnesium (23c.2)
26.2: Functions of magnesium (23c.2)
CCS
27.8: Copper chaperones (24b.7)
cellular energy
25.2: Functions of magnesium (23c.2)
26.2: Functions of magnesium (23c.2)
cellular_respiration
27.5: Cytochrome c oxidase (24b.4)
central nervous system
17.12: Ascorbic acid in other cells and fluids (19.12)
cerebrospinal fluid
17.12: Ascorbic acid in other cells and fluids (19.12)
ceruloplasmin
27.1: Copper (24b.0)
27.2: Serum copper (24b.1)
27.3: Serum ceruloplasmin (24b.2)
ceruloplasmin
27.8: Copper chaperones (24b.7)
chaperones
27.8: Copper chaperones (24b.7)
chemical analysis
12.2: Chemical analysis of cadavers (14.1)
child growth
11.3: Use of anthropometric indices in clinical settings (13.2)
child growth standards
9.3: Growth indices, indicators, and recommended growth reference data (10.2)
child nutrition
11.5: Applications of anthropometry in public health (13.4)
childhood growth
8.4: Interpretation and evaluation of anthropometric data
childhood obesity
9.3: Growth indices, indicators, and recommended growth reference data (10.2)
9.5: BMI in children and adolescents (10.4)
childhood_nutrition
28.4: Linear growth (24c.4)
Children
11: Evaluation of Anthropometric Data (Chapter 13)
11.4: Screening systems to identify individuals at risk in a population (13.3)
15.2: Indices of vitamin A status (18a.2)
22: Vitamin B12 (Chapter 22b)
23.5: Nutrient reference values for calcium (23a.5)
24.3: Interpretive criterion- serum phosphate (23b.3)
24.6: Nutrient reference values (23b.6)
28.13: Taste acuity (24c.13)
28.6: Hair zinc concentrations (24c.6)
8.1: Measurements, indices, and indicators
9.2: Measurements of body size (10.1)
9.5: BMI in children and adolescents (10.4)
children's health
11.2: Anthropometric indices and their modes of expression (13.1)
12.7: Densitometry (14.6)
cholecalciferol
16: Vitamin D (Chapter 18b)
16.2: Metabolism of vitamin D (18b.2)
chronic
25.7: Occurrence of dietary deficiency (23c.7)
chronic disease
1.3: Nutritional assessment methods
13.1: Biomarkers to assess nutritional status (15.1)
13.2: Biomarkers of exposure (15.2)
13.4: Biomarkers of function (15.4)
14.10: Multiple biomarkers (17.10)
14.2: Hemoglobin (17.2)
24: Phosphorus (Chapter 23b)
25.8: Indices of magnesium status (23c.8)
26: Magnesium (Chapter 23c)
26.8: Indices of magnesium status (23c.8)
3.1: Methods for measuring food consumption of individuals
3.2: Technical improvements in food consumption measurements
chronic disease prevention
5.1: Development of Nutrient Reference Values (8a.1)
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
5.5: United States and Canada- Dietary Reference Intakes (8a.5)
chronic disease risk
3.3: Selecting an appropriate method
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
chronic diseases
1: Introduction
1.1: New developments in nutritional assessment
13.7: Application of new technologies (15.7)
5.7: WHO/FAO Nutrient Reference Values (8a.7)
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)
chronic hypophosphatemia
24.9: Inherited and tumor-induced disorders of phosphate metabolism
chronic illness
1.2: Nutritional assessment systems
chronic inflammation
26.4: Magnesium deficiency in humans (23c.4)
chronic malnutrition
10.1: Anthropometric assessment of body composition (11.0)
chronicdisease
14.4: Red-cell indices
7.3: Examples of diet quality indices (8c.3)
chronicinflammation
7.8: Dietary Inflammatory Indices (8c.8)
circumference
10: Body composition (Chapter 11)
citrus
17.1: Vitamin C (19.1)
clinical
18.1: Thiamine (20a.1)
clinical application
12: Body Composition- Laboratory Methods (Chapter 14)
clinical applications
12.8: Total body electrical conductivity (14.7)
clinical assessment
12.12: Dual energy X-ray absorptiometry (14.11)
clinical assessments
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)
clinical conditions
28.5: Urine zinc (24c.5)
clinical diagnosis
16.10: Other vitamin D metabolites (18b.11)
clinical guidance
13.1: Biomarkers to assess nutritional status (15.1)
clinical improvement
18.2: Biochemical indices of thiamine status (20a.2)
clinical manifestations
24.8: Abnormalities in serum phosphate (23b.8)
clinical methods
25: Phosphorus (Chapter 23b)
clinical research
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
clinical tests
25.8: Indices of magnesium status (23c.8)
clinical trials
29: Selenium (Chapter 25b)
29.1: Overview (25b.1)
29.10: Questions Remain (25b.10)
clinical_signs
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
clinicalGuidelines
16.7: Interpretive criteria (18b.8)
clinicalmanifestations
17.4: Deficiency of vitamin C in humans (19.4)
clinicalMarkers
17: Vitamin C (Chapater 19)
clinicaltrials
1.5: The design of nutritional assessment systems
cobalamin
22.1: Introduction (22b.1)
coenzyme
18.1: Thiamine (20a.1)
coenzymes
19: Riboflavin (Chapter 20b)
19.1: Riboflavin (20b.1)

cognitive development
18: Thiamine (Chapter 20a)
cognitive function
28.1: Zinc (24c.1)
cognitive performance
13.4: Biomarkers of function (15.4)
cognitive process
4.2: Sources of measurement error
collagen
17.2: Functions of Vitamin C (19.2)
27.6: Skin lysyl oxidase (24b.5)
collagenCrosslinking
17.14: Functional tests of vitamin C status (19.14)
colon
26.3: Absorption and metabolism (23c.3)
complete blood count
14.5: Red-cell distribution width (17.5)
computerized tomography
12.10: Computerized tomography (14.9)
concentration
24.3: Interpretive criterion- serum phosphate (23b.3)
28.3: Serum zinc (24c.3)
concentrations
28.16: Leukocyte zinc (24c.16)
connective tissue
27.6: Skin lysyl oxidase (24b.5)
consumption
2: Food available for consumption at National and Household Levels
consumption surveys
2.3: Household Surveys
cookingEffects
17.5: Food sources and dietary intakes of vitamin C (19.5)
copper
27: Copper (Chapter 24b)
27.1: Copper (24b.0)
27.2: Serum copper (24b.1)
27.3: Serum ceruloplasmin (24b.2)
27.4: Erythrocyte superoxide dismutase (24b.3)
27.6: Skin lysyl oxidase (24b.5)
27.8: Copper chaperones (24b.7)
27.9: Other indices of copper status (24b.8)
copper deficiency
27.1: Copper (24b.0)
27.4: Erythrocyte superoxide dismutase (24b.3)
copper_deficiency
27.5: Cytochrome c oxidase (24b.4)
coronary artery disease
12.10: Computerized tomography (14.9)
CT
12.10: Computerized tomography (14.9)
cytochrome_c_oxidase
27.5: Cytochrome c oxidase (24b.4)

D

dairy products
25.5: Food sources and dietary Intakes
26.5: Food sources and dietary Intakes (23c.5)
dark chocolate
26.5: Food sources and dietary Intakes (23c.5)
DASH
7.6: Dietary Approaches to Stop Hypertension scores
data collection
4.5: Summary
data collection methods
2.3: Household Surveys
dataanalysis
1.5: The design of nutritional assessment systems
deamination
27.7: Diamine oxidase (24b.6)
deficiencies
16: Vitamin D (Chapter 18b)
16.4: Food sources and dietary intakes (18b.4)
19.1: Riboflavin (20b.1)
deficiency
13: Biomarkers (Chapter 15)
13.6: Evaluation of the selected nutritional biomarker (15.6)
14.1: Introduction and functions of iron (17.1)
16.10: Other vitamin D metabolites (18b.11)
16.11: Serum parathyroid hormone (18b.13)
16.12: Calcium and phosphorus in serum and urine (18b.14)
16.2: Metabolism of vitamin D (18b.2)
16.3: Vitamin D deficiency in humans (18b.3)
16.5: Nutrient reference values for vitamin D (18b.5)
16.7: Interpretive criteria (18b.8)
17: Vitamin C (Chapater 19)
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
17.14: Functional tests of vitamin C status (19.14)
17.8: Plasma ascorbic acid (19.8)
18.1: Thiamine (20a.1)
18.2: Biochemical indices of thiamine status (20a.2)
18.3: Conclusions (20a.3)
19: Riboflavin (Chapter 20b)
19.2: Biomarkers of riboflavin status (20b.2)
20: Niacin (Chapter 20c)
21.1: Folate (22a.1)
21.2: Biomarkers of folate status (Chapter 22a.2)
22.1: Introduction (22b.1)
23: Calcium (Chapter 23a)
25: Phosphorus (Chapter 23b)
25.4: Magnesium deficiency in humans
25.7: Occurrence of dietary deficiency (23c.7)
25.8: Indices of magnesium status (23c.8)
26: Magnesium (Chapter 23c)
26.1: Introduction (23c.1)
26.8: Indices of magnesium status (23c.8)
27: Copper (Chapter 24b)
27.2: Serum copper (24b.1)
27.6: Skin lysyl oxidase (24b.5)
27.8: Copper chaperones (24b.7)
27.9: Other indices of copper status (24b.8)
28: Zinc (Chapter 24c)
28.1: Zinc (24c.1)
28.18: Multiple biomarkers (24c.18)
28.3: Serum zinc (24c.3)
29.1: Overview (25b.1)
deficiencyassessment
22: Vitamin B12 (Chapter 22b)
deforestation
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)
dehydroascorbicacid
17.3: Absorption and metabolism of vitamin C (19.3)
densitometry
12: Body Composition- Laboratory Methods (Chapter 14)
12.7: Densitometry (14.6)
density
12.2: Chemical analysis of cadavers (14.1)
detection thresholds
28.13: Taste acuity (24c.13)
deuterium
12.4: Total body water from isotope dilution (14.3)
developing countries
2.3: Household Surveys
diabetes
10.1: Anthropometric assessment of body composition (11.0)
12.14: Summary - Method comparisons (14.13)
26.8: Indices of magnesium status (23c.8)
diagnosis
13.6: Evaluation of the selected nutritional biomarker (15.6)
diagnostic imaging
12.13: Ultrasound (14.12)
diagnostic information
14.10: Multiple biomarkers (17.10)
diagnostic tool
12.12: Dual energy X-ray absorptiometry (14.11)
diarrhea
26.6: Effects of high magnesium intakes (23c.6)
diet
13: Biomarkers (Chapter 15)
17.5: Food sources and dietary intakes of vitamin C (19.5)
2: Food available for consumption at National and Household Levels
3.2: Technical improvements in food consumption measurements
7.8: Dietary Inflammatory Indices (8c.8)
diet planning
6: Evaluation of intakes and diets (Chapter 8b)
diet quality
7: Dietary guidelines and quality (Chapter 8c)
7.12: Summary of diet quality indices (8c.12)
7.2: Assessing diet quality (8c.2)
7.9: Diet quality based on level of food processing (8c.9)
dietary
16.12: Calcium and phosphorus in serum and urine (18b.14)
16.4: Food sources and dietary intakes (18b.4)
17.6: Nutrient reference values for vitamin C (19.6)
19.1: Riboflavin (20b.1)
23.4: Dietary sources and intakes of calcium (23a.4)
25.7: Occurrence of dietary deficiency (23c.7)
27.9: Other indices of copper status (24b.8)
dietary allowance
29.4: Selenium Requirements (25b.4)
dietary assessment
1.3: Nutritional assessment methods
1.6: Important characteristics of assessment measures
13.1: Biomarkers to assess nutritional status (15.1)
13.2: Biomarkers of exposure (15.2)
3.1: Methods for measuring food consumption of individuals
3.2: Technical improvements in food consumption measurements
3.3: Selecting an appropriate method
4.2: Sources of measurement error
4.3: Implications of measurement error for estimated dietary intakes
4.4: Minimizing measurement error through data collection procedures
6.4: Using the NRVs for planning nutritious diets (8b.4)

dietary assessment methods
- 4.1: Measurement error in dietary intake data
- 6: Evaluation of intakes and diets (Chapter 8b)

dietary behaviors
- 1.1: New developments in nutritional assessment

dietary biomarkers
- 1.2: Nutritional assessment systems
- 13.2: Biomarkers of exposure (15.2)

dietary components
- 4.3: Implications of measurement error for estimated dietary intakes

dietary copper
- 27.6: Skin lysyl oxidase (24b.5)

dietary counseling
- 6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)
- 6.4: Using the NRVs for planning nutritious diets (8b.4)

dietary data
- 2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
- 4.5: Summary
- 6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

dietary diversity
- 1.4: Nutritional assessment indices and indicators

dietary diversity score
- 2.1: Introduction

dietary factors
- 13.7: Application of new technologies (15.7)
- 2.5: National and household surveys in the global assessment of food security and diet quality by age and sex

dietary fiber
- 5.6: European Dietary Reference Values (8a.6)

dietary guidance
- 7.1: Food-based dietary guidelines (8c.1)

dietary guidelines
- 4.2: Sources of measurement error
- 5.5: United States and Canada- Dietary Reference Intakes (8a.5)
- 6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)
- 6.4: Using the NRVs for planning nutritious diets (8b.4)
- 7: Dietary guidelines and quality (Chapter 8c)
- 7.12: Summary of diet quality indices (8c.12)
- 7.2: Assessing diet quality (8c.2)
- 7.4: Alternate Healthy Eating Index (8c.4)
- 7.7: Healthy Diet Indicator (8c.7)

dietary history
- 3.1: Methods for measuring food consumption of individuals
- 3.3: Selecting an appropriate method

dietary imbalances
- 8.2: Advantages and limitations of anthropometry

dietary instruments
- 3.3: Selecting an appropriate method

dietary intake
- 13.3: Biomarkers of status (15.3)
- 17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
- 17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
- 17.8: Plasma ascorbic acid (19.8)
- 19: Riboflavin (Chapter 20b)
- 19.2: Biomarkers of riboflavin status (20b.2)
- 20: Niacin (Chapter 20c)
- 20.1: Niacin - Introduction (20c.1)
- 20.2: Indices of niacin status (20c.2)
- 21.2: Biomarkers of folate status (Chapter 22a.2)
- 22.2: Biomarkers of vitamin B12 status (22b.2)
- 23.3: Absorption and metabolism of calcium (23a.3)
- 23.5: Nutrient reference values for calcium (23a.5)
- 24: Phosphorus (Chapter 23b)
- 24.4: Phosphate balance (23b.4)
- 24.6: Nutrient reference values (23b.6)
- 24.8: Abnormalities in serum phosphate (23b.8)
- 25: Phosphorus (Chapter 23b)
- 25.3: Absorption and metabolism (23c.3)
- 25.8: Indices of magnesium status (23c.8)
- 26: Magnesium (Chapter 23c)
- 26.3: Absorption and metabolism (23c.3)
- 26.8: Indices of magnesium status (23c.8)
- 27.2: Serum copper (24b.1)
- 28.1: Zinc (24c.1)
- 28.18: Multiple biomarkers (24c.18)
- 28.5: Urine zinc (24c.5)
- 3.2: Technical improvements in food consumption measurements
- 4: Measurement Error in Dietary assessment (Chapter 5)
- 4.1: Measurement error in dietary intake data
- 4.2: Sources of measurement error
- 5.7: WHO/FAO Nutrient Reference Values (8a.7)
- 7.12: Summary of diet quality indices (8c.12)
- 7.9: Diet quality based on level of food processing (8c.9)

dietary intakes
- 17.12: Ascorbic acid in other cells and fluids (19.12)
- 28.2: Dietary Zinc Intake (24c.2)

dietary iron
- 14.7: Serum ferritin (17.7)

dietary methods
- 13.5: Factors affecting choice of nutritional biomarkers (15.5)

dietary pattern
- 5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)

dietary patterns
- 23.6: Calcium deficiency (23a.6)
- 28.2: Dietary Zinc Intake (24c.2)
- 5: Nutrient reference values (Chapter 8a)
- 5.1: Development of Nutrient Reference Values (8a.1)
- 7: Dietary guidelines and quality (Chapter 8c)
- 7.2: Assessing diet quality (8c.2)
- 7.6: Dietary Approaches to Stop Hypertension scores

dietary quality
- 2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)

dietary recommendations
- 29.4: Selenium Requirements (25b.4)
- 5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
- 7.7: Healthy Diet Indicator (8c.7)

dietary reference intakes
- 26.7: Occurrence of dietary deficiency (23c.7)
- 5.1: Development of Nutrient Reference Values (8a.1)
- 5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
- 5.5: United States and Canada- Dietary Reference Intakes (8a.5)

dietary reference values
- 5.4: U.K. Dietary Reference Values (8a.4)
- 5.6: European Dietary Reference Values (8a.6)

dietary requirements
- 26.7: Occurrence of dietary deficiency (23c.7)
- 5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

dietary research
- 4.2: Sources of measurement error

dietary Se
- 29.7: Metabolism (25b.7)

dietary supplements
- 26.6: Effects of high magnesium intakes (23c.6)
- 29.10: Questions Remain (25b.10)
- 4.5: Summary

dietary survey
- 1: Introduction

dietary surveys
- 2.3: Household Surveys

dietary zinc
- 28: Zinc (Chapter 24c)
- 28.11: Oxidative stress and DNA integrity (24c.11)
- 28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
- 28.6: Hair zinc concentrations (24c.6)
- 28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)

dietaryadequacy
- 7.3: Examples of diet quality indices (8c.3)

dietaryassessment
- 6.3: Using NRVs to assess nutrient intakes of groups (8b.3)
- 7.3: Examples of diet quality indices (8c.3)

dietarydeficiency
- 22: Vitamin B12 (Chapter 22b)

dietaryenergy
- 2.2: Food Balance Sheets

dietaryfactors
- 7.8: Dietary Inflammatory Indices (8c.8)

dietaryhabits
- 7.10: Lower-burden diet quality indices (8c.10)

dietaryIntake
- 16.5: Nutrient reference values for vitamin D (18b.5)
- 17: Vitamin C (Chapater 19)
- 7.10: Lower-burden diet quality indices (8c.10)

dietaryiron
- 14: Iron (Chapter 17)

dietarypatterns
- 7.10: Lower-burden diet quality indices (8c.10)
- 7.5: Mediterranean Diet Score or Index (8c.5)

dietaryrecommendations
- 16.5: Nutrient reference values for vitamin D (18b.5)
- 6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

dietaryrequirements
- 22: Vitamin B12 (Chapter 22b)

dietquality
7.10: Lower-burden diet quality indices (8c.10)
7.3: Examples of diet quality indices (8c.3)
diets
6.4: Using the NRVs for planning nutritious diets (8b.4)
differentiation
16.1: Functions of vitamin D (18b.1)
Digestion
29.6: Enteric Absorption (25b.6)
diketo form
17.1: Vitamin C (19.1)
disease
13.5: Factors affecting choice of nutritional biomarkers (15.5)
25.4: Magnesium deficiency in humans
disease prevention
13.1: Biomarkers to assess nutritional status (15.1)
diseases
12.14: Summary - Method comparisons (14.13)
disorders
18.1: Thiamine (20a.1)
distribution
23.1: Calcium (23a.1)
DNA
17.14: Functional tests of vitamin C status (19.14)
26.2: Functions of magnesium (23c.2)
DNA biosynthesis
21: Folate (Chapter 22a)
DNA damage
28.11: Oxidative stress and DNA integrity (24c.11)
DNA repair
21.3: Conclusions (22a.3)
DNA synthesis
21.3: Conclusions (22a.3)
25.2: Functions of magnesium (23c.2)
DRI
26.7: Occurrence of dietary deficiency (23c.7)
DXA
12.12: Dual energy X-ray absorptiometry (14.11)
23.9: Measurement of bone mineral content and bone mineral density (23a.9)
dyslipidemia
26.4: Magnesium deficiency in humans (23c.4)

E

EAR
23.5: Nutrient reference values for calcium (23a.5)
26.7: Occurrence of dietary deficiency (23c.7)
eating patterns
1.3: Nutritional assessment methods
EFSA
17.6: Nutrient reference values for vitamin C (19.6)
Elderly
10.3: Assessment of fat-free mass (11.2)
16.10: Other vitamin D metabolites (18b.11)
17: Vitamin C (Chapater 19)
17.4: Deficiency of vitamin C in humans (19.4)
22: Vitamin B12 (Chapter 22b)
elderly subjects
12.8: Total body electrical conductivity (14.7)
electrical conductivity
12.8: Total body electrical conductivity (14.7)
electrolyte imbalance
24.8: Abnormalities in serum phosphate (23b.8)
electrolytes
25.6: Effects of high magnesium intakes (23c.6)
electron_transport_chain
27.5: Cytochrome c oxidase (24b.4)
electrondonation
17.2: Functions of Vitamin C (19.2)
endocrine
16.1: Functions of vitamin D (18b.1)
endocrine hormones
24: Phosphorus (Chapter 23b)
energy
5: Nutrient reference values (Chapter 8a)
5.6: European Dietary Reference Values (8a.6)
energy balance
10.2: Assessment of body fat (11.1)
energy expenditure
8.2: Advantages and limitations of anthropometry
energy intake
4.3: Implications of measurement error for estimated dietary intakes
6.4: Using the NRVs for planning nutritious diets (8b.4)
energy metabolism
18: Thiamine (Chapter 20a)
20.1: Niacin - Introduction (20c.1)
energy production
19: Riboflavin (Chapter 20b)
19.1: Riboflavin (20b.1)
energy requirement
6: Evaluation of intakes and diets (Chapter 8b)
energy requirements
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
5.4: U.K. Dietary Reference Values (8a.4)
enolic form
17.1: Vitamin C (19.1)
environmentally sustainable
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)
enzymatic function
27.1: Copper (24b.9)
enzymatic reactions
25.2: Functions of magnesium (23c.2)
26.2: Functions of magnesium (23c.2)
enzyme
27.6: Skin lysyl oxidase (24b.5)
enzyme interaction
26.2: Functions of magnesium (23c.2)
enzymes
27: Copper (Chapter 24b)
27.3: Serum ceruloplasmin (24b.2)
epidemiologic research
4: Measurement Error in Dietary assessment (Chapter 5)
EpidemiologicalStudies
7.5: Mediterranean Diet Score or Index (8c.5)
epidemiology
1.5: The design of nutritional assessment systems
29.1: Overview (25b.1)
4.1: Measurement error in dietary intake data
7.2: Assessing diet quality (8c.2)
9.4: Body mass index in adults (10.3)
9.5: BMI in children and adolescents (10.4)
epigeneticMarks
17.14: Functional tests of vitamin C status (19.14)
ergocalciferol
16.2: Metabolism of vitamin D (18b.2)
erythrocyte
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
19.2: Biomarkers of riboflavin status (20b.2)
erythrocytes
14.8: Erythrocyte protoporphyrin
18.2: Biochemical indices of thiamine status (20a.2)
27.4: Erythrocyte superoxide dismutase (24b.3)
27.5: Cytochrome c oxidase (24b.4)
27.8: Copper chaperones (24b.7)
27.9: Other indices of copper status (24b.8)
28.15: Erythrocyte zinc (24c.15)
28.16: Leukocyte zinc (24c.16)
erythropoiesis
14: Iron (Chapter 17)
14.11: Emerging iron status indicators (17.11)
14.4: Red-cell indices
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
14.9: Serum soluble transferrin receptor (17.9)
essential functions
25.1: Introduction (23c.1)
26.1: Introduction (23c.1)
essential nutrients
24.1: Phosphorus (23b.1)
estimated average requirement
26.7: Occurrence of dietary deficiency (23c.7)
5.4: U.K. Dietary Reference Values (8a.4)
ethicalissues
1.5: The design of nutritional assessment systems
ethnicity
10.2: Assessment of body fat (11.1)
eukaryota
29.8: Biochemical Functions (25b.8)
evaluation
7: Dietary guidelines and quality (Chapter 8c)
evidence synthesis
7.1: Food-based dietary guidelines (8c.1)
excess
28.18: Multiple biomarkers (24c.18)
excessive intake
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)
excretion
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
28.5: Urine zinc (24c.5)
29.7: Metabolism (25b.7)
expenditure
2: Food available for consumption at National and Household Levels
expertGroups
17.6: Nutrient reference values for vitamin C (19.6)

exposure
13: Biomarkers (Chapter 15)
expression
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
extracellular
24.4: Phosphate balance (23b.4)
25.1: Introduction (23c.1)
26.1: Introduction (23c.1)

F

FAO
5.7: WHO/FAO Nutrient Reference Values (8a.7)
7.1: Food-based dietary guidelines (8c.1)
fat free weight
12.2: Chemical analysis of cadavers (14.1)
fat mass
10.1: Anthropometric assessment of body composition (11.0)
fatfreemass
10: Body composition (Chapter 11)
fattyfish
16.4: Food sources and dietary intakes (18b.4)
FBS
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
females
25.7: Occurrence of dietary deficiency (23c.7)
ferritin
14: Iron (Chapter 17)
14.1: Introduction and functions of iron (17.1)
14.11: Emerging iron status indicators (17.11)
14.4: Red-cell indices
14.7: Serum ferritin (17.7)
14.8: Erythrocyte protoporphyrin
ferrochetalase
14.8: Erythrocyte protoporphyrin
fertilizers
24.1: Phosphorus (23b.1)
fetal growth
11.1: Introduction (13.0)
FGF23
24.5: Hormonal regulation of serum phosphate (23b.5)
FISH
7.4: Alternate Healthy Eating Index (8c.4)
fitness
12.9: Bioelectrical impedance (14.8)
folate
13.4: Biomarkers of function (15.4)
21: Folate (Chapter 22a)
21.1: Folate (22a.1)
21.2: Biomarkers of folate status (Chapter 22a.2)
21.3: Conclusions (22a.3)
5: Nutrient reference values (Chapter 8a)
folate deficiency
14.5: Red-cell distribution width (17.5)
21.3: Conclusions (22a.3)
folic acid
21.1: Folate (22a.1)
Food
2: Food available for consumption at National and Household Levels
23.4: Dietary sources and intakes of calcium (23a.4)
29.6: Enteric Absorption (25b.6)
7: Dietary guidelines and quality (Chapter 8c)
food analysis
4.1: Measurement error in dietary intake data
food and nutrition
13.1: Biomarkers to assess nutritional status (15.1)
food availability
2.1: Introduction
food balance sheets
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
food chain
29.5: Food Sources (25b.5)
food classification
7.9: Diet quality based on level of food processing (8c.9)
food composition
7.12: Summary of diet quality indices (8c.12)
food consumption
2.1: Introduction
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
3: Food consumption of individuals
3.1: Methods for measuring food consumption of individuals
3.2: Technical improvements in food consumption measurements
4: Measurement Error in Dietary assessment (Chapter 5)
4.2: Sources of measurement error
4.5: Summary
food consumption patterns
2.3: Household Surveys
food frequency questionnaire
3.1: Methods for measuring food consumption of individuals
3.3: Selecting an appropriate method
food habits
3.3: Selecting an appropriate method
food insecurity
2.1: Introduction
food loss
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
food policy
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
food quality
2.1: Introduction
food record
28.2: Dietary Zinc Intake (24c.2)
food regulation
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
food safety
24.6: Nutrient reference values (23b.6)
7.2: Assessing diet quality (8c.2)
Food Security
11.5: Applications of anthropometry in public health (13.4)
2.3: Household Surveys
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
7.1: Food-based dietary guidelines (8c.1)
food systems
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)
food waste
2.3: Household Surveys
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
foodbalance
2.2: Food Balance Sheets
foodconsumption
2.2: Food Balance Sheets
foodgroupdiversity
7.10: Lower-burden diet quality indices (8c.10)
foodloss
2.2: Food Balance Sheets
foodproduction
2.2: Food Balance Sheets
foods
25.5: Food sources and dietary Intakes
foodsecurity
2.2: Food Balance Sheets
foodstatistics
2.2: Food Balance Sheets
foodsupply
2.2: Food Balance Sheets
fortification
16.4: Food sources and dietary intakes (18b.4)
23.4: Dietary sources and intakes of calcium (23a.4)
fracture risk
23.6: Calcium deficiency (23a.6)
fracture_risk
16.11: Serum parathyroid hormone (18b.13)
fractures
16.3: Vitamin D deficiency in humans (18b.3)
frailty
12.5: Multiple dilution methods (14.4)
freshFruits
17.5: Food sources and dietary intakes of vitamin C (19.5)
fruits
7.5: Mediterranean Diet Score or Index (8c.5)
7.6: Dietary Approaches to Stop Hypertension scores
7.7: Healthy Diet Indicator (8c.7)

G

genetic
13.5: Factors affecting choice of nutritional biomarkers (15.5)
genetic disorders
24.9: Inherited and tumor-induced disorders of phosphate metabolism
genetic mutations
17.1: Vitamin C (19.1)
genetic variations
13.7: Application of new technologies (15.7)
genetic_defects
27.5: Cytochrome c oxidase (24b.4)
geographic variation
29.5: Food Sources (25b.5)
gingivalbleeding
17.4: Deficiency of vitamin C in humans (19.4)
global
2: Food available for consumption at National and Household Levels

global databases
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
global use
7.12: Summary of diet quality indices (8c.12)
glutathione
19.2: Biomarkers of riboflavin status (20b.2)
glutathione peroxidase
29: Selenium (Chapter 25b)
Glutathione Peroxidase 3
29.4: Selenium Requirements (25b.4)
Glutathione Peroxidases
29.8: Biochemical Functions (25b.8)
glycine metabolism
21: Folate (Chapter 22a)
glycolysis
25.2: Functions of magnesium (23c.2)
26.2: Functions of magnesium (23c.2)
granulocytes
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
greenhouse gas emissions
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)
growth
11: Evaluation of Anthropometric Data (Chapter 13)
11.5: Applications of anthropometry in public health (13.4)
28: Zinc (Chapter 24c)
9.2: Measurements of body size (10.1)
growth assessment
11.3: Use of anthropometric indices in clinical settings (13.2)
growth charts
11.3: Use of anthropometric indices in clinical settings (13.2)
9.5: BMI in children and adolescents (10.4)
growth limiting factor
24.1: Phosphorus (23b.1)
growth measurements
11.3: Use of anthropometric indices in clinical settings (13.2)
growth monitoring
8.2: Advantages and limitations of anthropometry
8.3: Errors in anthropometry
growth reference
9: Body size (Chapter 10)
growth standards
1.7: Evaluation of nutritional assessment indices
11: Evaluation of Anthropometric Data (Chapter 13)
11.1: Introduction (13.0)
11.2: Anthropometric indices and their modes of expression (13.1)
11.4: Screening systems to identify individuals at risk in a population (13.3)
8: Introduction to Anthropometry (Chapter 9)
8.4: Interpretation and evaluation of anthropometric data
growth_standards
28.4: Linear growth (24c.4)
gustin
28.17: Salivary zinc (24c.17)

H

hair
27.9: Other indices of copper status (24b.8)
hair growth
28.6: Hair zinc concentrations (24c.6)
hair zinc
28.7: Nail zinc (24c.7)
harmonization
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
Hb concentration
14.3: Hematocrit or packed cell volume
HCES
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
HDI
7.7: Healthy Diet Indicator (8c.7)
head circumference
9.2: Measurements of body size (10.1)
health
12.9: Bioelectrical impedance (14.8)
13: Biomarkers (Chapter 15)
13.1: Biomarkers to assess nutritional status (15.1)
13.5: Factors affecting choice of nutritional biomarkers (15.5)
16.4: Food sources and dietary intakes (18b.4)
17.6: Nutrient reference values for vitamin C (19.6)
17.7: Indices of vitamin C status (19.7)
19: Riboflavin (Chapter 20b)
19.1: Riboflavin (20b.1)
19.2: Biomarkers of riboflavin status (20b.2)
20.1: Niacin - Introduction (20c.1)
22.1: Introduction (22b.1)
23.7: Effects of high intakes of calcium (23a.7)
24.1: Phosphorus (23b.1)
25: Phosphorus (Chapter 23b)
25.1: Introduction (23c.1)
26.1: Introduction (23c.1)
26.7: Occurrence of dietary deficiency (23c.7)
27: Copper (Chapter 24b)
28.1: Zinc (24c.1)
28.17: Salivary zinc (24c.17)
28.3: Serum zinc (24c.3)
29.1: Overview (25b.1)
29.10: Questions Remain (25b.10)
29.4: Selenium Requirements (25b.4)
29.9: Assessing Selenium Status (25b.9)
5.4: U.K. Dietary Reference Values (8a.4)
5.7: WHO/FAO Nutrient Reference Values (8a.7)
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
7: Dietary guidelines and quality (Chapter 8c)
7.2: Assessing diet quality (8c.2)
7.3: Examples of diet quality indices (8c.3)
7.8: Dietary Inflammatory Indices (8c.8)
9.5: BMI in children and adolescents (10.4)
health applications
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)
health assessment
1.7: Evaluation of nutritional assessment indices
10.2: Assessment of body fat (11.1)
12.12: Dual energy X-ray absorptiometry (14.11)
22.2: Biomarkers of vitamin B12 status (22b.2)
8: Introduction to Anthropometry (Chapter 9)
health assessments
8.4: Interpretation and evaluation of anthropometric data
health benefits
20.2: Indices of niacin status (20c.2)
health care providers
11.3: Use of anthropometric indices in clinical settings (13.2)
health effects
25.6: Effects of high magnesium intakes (23c.6)
health implications
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
health indicators
14.3: Hematocrit or packed cell volume
health interventions
1.3: Nutritional assessment methods
11.4: Screening systems to identify individuals at risk in a population (13.3)
health metrics
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
health nutrition
17.8: Plasma ascorbic acid (19.8)
health optimization
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
5.5: United States and Canada- Dietary Reference Intakes (8a.5)
health organization
1: Introduction
health outcomes
1.7: Evaluation of nutritional assessment indices
10.1: Anthropometric assessment of body composition (11.0)
10.3: Assessment of fat-free mass (11.2)
13.7: Application of new technologies (15.7)
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
7: Dietary guidelines and quality (Chapter 8c)
7.9: Diet quality based on level of food processing (8c.9)
health parameters
3.1: Methods for measuring food consumption of individuals
health policy
1.3: Nutritional assessment methods
health research
4.5: Summary
health risk
26: Magnesium (Chapter 23c)
health risks
9.4: Body mass index in adults (10.3)
health status
13.6: Evaluation of the selected nutritional biomarker (15.6)
healthassessment
16.6: Biomarkers of vitamin D status (18b.6)
healthbenefits
17.3: Absorption and metabolism of vitamin C (19.3)
healthcare
16.7: Interpretive criteria (18b.8)
healthcare interventions
1.2: Nutritional assessment systems
Healthfulness
7.5: Mediterranean Diet Score or Index (8c.5)

healthindicators
7.10: Lower-burden diet quality indices (8c.10)
healthinterventions
18.3: Conclusions (20a.3)
healthrisks
16.5: Nutrient reference values for vitamin D (18b.5)
healthy
25.7: Occurrence of dietary deficiency (23c.7)
healthy diet indicator
7.7: Healthy Diet Indicator (8c.7)
healthy diets
7.1: Food-based dietary guidelines (8c.1)
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)
healthy eating
7.3: Examples of diet quality indices (8c.3)
Healthy Eating Index
7.4: Alternate Healthy Eating Index (8c.4)
healthy study
12.2: Chemical analysis of cadavers (14.1)
heart arrhythmias
26.6: Effects of high magnesium intakes (23c.6)
heart disease
29.1: Overview (25b.1)
heart failure
26.4: Magnesium deficiency in humans (23c.4)
HEI
7.3: Examples of diet quality indices (8c.3)
height
9.1: Anthropometric assessment of body size (10.0)
9.2: Measurements of body size (10.1)
hematocrit
14.3: Hematocrit or packed cell volume
hematology
14.9: Serum soluble transferrin receptor (17.9)
heme synthesis
14.8: Erythrocyte protoporphyrin
hemoglobin
13.6: Evaluation of the selected nutritional biomarker (15.6)
14.1: Introduction and functions of iron (17.1)
14.2: Hemoglobin (17.2)
14.4: Red-cell indices
hemolysis
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
hemolyzed
24.3: Interpretive criterion- serum phosphate (23b.3)
hemorrhages
17.4: Deficiency of vitamin C in humans (19.4)
heparin
24.3: Interpretive criterion- serum phosphate (23b.3)
hepcidin
14.11: Emerging iron status indicators (17.11)
14.7: Serum ferritin (17.7)
hereditary hypophosphatemic rickets
24.9: Inherited and tumor-induced disorders of phosphate metabolism
high performance liquid chromatography
18.2: Biochemical indices of thiamine status (20a.2)
histamine
27.7: Diamine oxidase (24b.6)
Homeostasis
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)
23.8: Biochemical indices of calcium status (23a.8)
25.3: Absorption and metabolism (23c.3)
26.3: Absorption and metabolism (23c.3)
28.18: Multiple biomarkers (24c.18)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)
homocysteine
21.2: Biomarkers of folate status (Chapter 22a.2)
22.2: Biomarkers of vitamin B12 status (22b.2)
hormonal responses
23.6: Calcium deficiency (23a.6)
hormone function
23.2: Functions of calcium (23a.2)
hormone_regulation
24.5: Hormonal regulation of serum phosphate (23b.5)
hormones
23: Calcium (Chapter 23a)
household consumption
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)
household surveys
2.1: Introduction
2.3: Household Surveys
HPLC
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
human body
24.2: Biological forms of phosphate and their measurement (23b.2)
26.1: Introduction (23c.1)
27.1: Copper (24b.0)
human diets
29.5: Food Sources (25b.5)
human health
29.3: Health Effects (25b.3)
human studies
28.11: Oxidative stress and DNA integrity (24c.11)
hydration
12.3: Total body potassium (TBK) (14.2)
12.4: Total body water from isotope dilution (14.3)
12.9: Bioelectrical impedance (14.8)
hydration status
12.8: Total body electrical conductivity (14.7)
hydrogen atoms
12.11: Magnetic resonance imaging (14.10)
hydrostatic weighing
12.7: Densitometry (14.6)
hydroxyapatite
23.2: Functions of calcium (23a.2)
24.2: Biological forms of phosphate and their measurement (23b.2)
hypercalcemia
23.8: Biochemical indices of calcium status (23a.8)
hypercupremia
27.2: Serum copper (24b.1)
hypermagnesemia
25.6: Effects of high magnesium intakes (23c.6)
26.6: Effects of high magnesium intakes (23c.6)
hyperphosphate
24.5: Hormonal regulation of serum phosphate (23b.5)
hyperphosphatemia
24: Phosphorus (Chapter 23b)
24.8: Abnormalities in serum phosphate (23b.8)
hypertension
25.8: Indices of magnesium status (23c.8)
26.4: Magnesium deficiency in humans (23c.4)
26.8: Indices of magnesium status (23c.8)
hypervitaminosis
16.12: Calcium and phosphorus in serum and urine (18b.14)
hyperzincuria
28.5: Urine zinc (24c.5)
hypocalcemia
23.8: Biochemical indices of calcium status (23a.8)
hypochromic
14: Iron (Chapter 17)
hypogeusia
28.13: Taste acuity (24c.13)
hypophosphatemia
24: Phosphorus (Chapter 23b)
24.8: Abnormalities in serum phosphate (23b.8)

I

Imaging
12.12: Dual energy X-ray absorptiometry (14.11)
imaging technique
12.10: Computerized tomography (14.9)
12.11: Magnetic resonance imaging (14.10)
immune cells
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
immune function
15.1: Introduction (18a.1)
28: Zinc (Chapter 24c)
28.1: Zinc (24c.1)
28.15: Erythrocyte zinc (24c.15)
28.18: Multiple biomarkers (24c.18)
immuneResponse
16.1: Functions of vitamin D (18b.1)
16.2: Metabolism of vitamin D (18b.2)
immunoassay
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
impedance
12.9: Bioelectrical impedance (14.8)
in vivo methods
12: Body Composition- Laboratory Methods (Chapter 14)
in vivo neutron activation analysis
12.6: In vivo activation analysis (14.5)
indicators
1.4: Nutritional assessment indices and indicators
indirect assessment
12: Body Composition- Laboratory Methods (Chapter 14)
individual assessment
6: Evaluation of intakes and diets (Chapter 8b)
individual consumption
3: Food consumption of individuals

infants
16.3: Vitamin D deficiency in humans (18b.3)
23.5: Nutrient reference values for calcium (23a.5)
24.6: Nutrient reference values (23b.6)
9.2: Measurements of body size (10.1)

infection
17.8: Plasma ascorbic acid (19.8)
27: Copper (Chapter 24b)

inflammation
13.1: Biomarkers to assess nutritional status (15.1)
13.5: Factors affecting choice of nutritional biomarkers (15.5)
14.11: Emerging iron status indicators (17.11)
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
14.7: Serum ferritin (17.7)
14.9: Serum soluble transferrin receptor (17.9)
25.4: Magnesium deficiency in humans
27.3: Serum ceruloplasmin (24b.2)
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)
7.8: Dietary Inflammatory Indices (8c.8)

informedconsent
1.5: The design of nutritional assessment systems

inorganic phosphate
24.7: Dietary sources

inorganic salts
24.2: Biological forms of phosphate and their measurement (23b.2)

Institute of Medicine
29.4: Selenium Requirements (25b.4)

insufficiency
16.7: Interpretive criteria (18b.8)

intake
25.5: Food sources and dietary Intakes
25.7: Occurrence of dietary deficiency (23c.7)

intake levels
25.6: Effects of high magnesium intakes (23c.6)

international harmonization
5.1: Development of Nutrient Reference Values (8a.1)

intervention
11.4: Screening systems to identify individuals at risk in a population (13.3)
21.2: Biomarkers of folate status (Chapter 22a.2)
28.18: Multiple biomarkers (24c.18)

intervention research
4: Measurement Error in Dietary assessment (Chapter 5)

interventions
11.5: Applications of anthropometry in public health (13.4)

intestinal absorption
23.3: Absorption and metabolism of calcium (23a.3)
24.4: Phosphate balance (23b.4)

intestinalhealth
17.3: Absorption and metabolism of vitamin C (19.3)

intestine
24.5: Hormonal regulation of serum phosphate (23b.5)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)

intoxication
16.12: Calcium and phosphorus in serum and urine (18b.14)

intracellular
25.1: Introduction (23c.1)

investigations
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)

Iodothyronine Deiodinases
29.8: Biochemical Functions (25b.8)

IOM
17.6: Nutrient reference values for vitamin C (19.6)

ion channels
25.2: Functions of magnesium (23c.2)

iron
13.2: Biomarkers of exposure (15.2)
14: Iron (Chapter 17)
14.1: Introduction and functions of iron (17.1)
14.11: Emerging iron status indicators (17.11)
14.2: Hemoglobin (17.2)
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
14.7: Serum ferritin (17.7)
14.8: Erythrocyte protoporphyrin
5: Nutrient reference values (Chapter 8a)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

iron absorption
23.7: Effects of high intakes of calcium (23a.7)

iron deficiency
14.11: Emerging iron status indicators (17.11)
14.2: Hemoglobin (17.2)
14.3: Hematocrit or packed cell volume
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
14.7: Serum ferritin (17.7)
14.8: Erythrocyte protoporphyrin

iron deficiency anemia
14.10: Multiple biomarkers (17.10)
14.9: Serum soluble transferrin receptor (17.9)

iron overload
14.11: Emerging iron status indicators (17.11)

iron status
14.10: Multiple biomarkers (17.10)

iron transport
14.9: Serum soluble transferrin receptor (17.9)

ironabsorption
17.2: Functions of Vitamin C (19.2)

irondeficiency
14: Iron (Chapter 17)
14.4: Red-cell indices

ischemic stroke
26.4: Magnesium deficiency in humans (23c.4)

isotope dilution
12.5: Multiple dilution methods (14.4)
17.13: Body pool size (19.13)

Isotopes
12.3: Total body potassium (TBK) (14.2)
12.4: Total body water from isotope dilution (14.3)

isotopic tracer
12.4: Total body water from isotope dilution (14.3)
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)

J

jejunum
26.3: Absorption and metabolism (23c.3)

K

Keshan disease
29.3: Health Effects (25b.3)

kidney
25.3: Absorption and metabolism (23c.3)
26.3: Absorption and metabolism (23c.3)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)

kidney dysfunction
25.6: Effects of high magnesium intakes (23c.6)
26.6: Effects of high magnesium intakes (23c.6)

kidney function
23.8: Biochemical indices of calcium status (23a.8)

kidney stones
23.7: Effects of high intakes of calcium (23a.7)

kidneyfunction
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)

kidneys
24.4: Phosphate balance (23b.4)
24.5: Hormonal regulation of serum phosphate (23b.5)
26.6: Effects of high magnesium intakes (23c.6)

L

laboratory
18.3: Conclusions (20a.3)

laboratory techniques
12.1: An introduction to techniques used to measure body composition (14.1)

lactatingwomen
22: Vitamin B12 (Chapter 22b)

lactation
23.3: Absorption and metabolism of calcium (23a.3)
24.6: Nutrient reference values (23b.6)

land conversion
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)

leafy vegetables
26.5: Food sources and dietary Intakes (23c.5)

legumes
7.5: Mediterranean Diet Score or Index (8c.5)

leucocytes
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)

leukocyte
17.7: Indices of vitamin C status (19.7)

leukocytes
17.12: Ascorbic acid in other cells and fluids (19.12)
17.14: Functional tests of vitamin C status (19.14)
27.5: Cytochrome c oxidase (24b.4)
28.16: Leukocyte zinc (24c.16)

lifespan
28.16: Leukocyte zinc (24c.16)

lifestylefactors
16.6: Biomarkers of vitamin D status (18b.6)

linear_growth
28.4: Linear growth (24c.4)

liquid chromatography
12.5: Multiple dilution methods (14.4)
liver
15.2: Indices of vitamin A status (18a.2)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)
livestock
29.3: Health Effects (25b.3)
low income countries
1.1: New developments in nutritional assessment
lymphocytes
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
28.16: Leukocyte zinc (24c.16)
lysyl oxidase
27.6: Skin lysyl oxidase (24b.5)

M

macrocephalic
9.2: Measurements of body size (10.1)
macrocytic
14.4: Red-cell indices
macronutrient distribution
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
Macronutrients
5.5: United States and Canada- Dietary Reference Intakes (8a.5)
magnesium
23.1: Calcium (23a.1)
25: Phosphorus (Chapter 23b)
25.1: Introduction (23c.1)
25.2: Functions of magnesium (23c.2)
25.3: Absorption and metabolism (23c.3)
25.4: Magnesium deficiency in humans
25.5: Food sources and dietary Intakes
25.6: Effects of high magnesium intakes (23c.6)
25.7: Occurrence of dietary deficiency (23c.7)
25.8: Indices of magnesium status (23c.8)
26: Magnesium (Chapter 23c)
26.2: Functions of magnesium (23c.2)
26.3: Absorption and metabolism (23c.3)
26.5: Food sources and dietary Intakes (23c.5)
26.6: Effects of high magnesium intakes (23c.6)
26.7: Occurrence of dietary deficiency (23c.7)
26.8: Indices of magnesium status (23c.8)
magnesium
26.1: Introduction (23c.1)
magnesium deficiency
26.4: Magnesium deficiency in humans (23c.4)
magnesium load test
25: Phosphorus (Chapter 23b)
magnetic field
12.11: Magnetic resonance imaging (14.10)
magnetic resonance imaging
12.11: Magnetic resonance imaging (14.10)
males
25.7: Occurrence of dietary deficiency (23c.7)
malnourished patients
12.5: Multiple dilution methods (14.4)
malnutrition
1: Introduction
1.2: Nutritional assessment systems
1.3: Nutritional assessment methods
1.5: The design of nutritional assessment systems
1.7: Evaluation of nutritional assessment indices
10: Body composition (Chapter 11)
10.3: Assessment of fat-free mass (11.2)
11: Evaluation of Anthropometric Data (Chapter 13)
11.1: Introduction (13.0)
11.2: Anthropometric indices and their modes of expression (13.1)
11.3: Use of anthropometric indices in clinical settings (13.2)
11.4: Screening systems to identify individuals at risk in a population (13.3)
11.5: Applications of anthropometry in public health (13.4)
15.2: Indices of vitamin A status (18a.2)
27.4: Erythrocyte superoxide dismutase (24b.3)
28.6: Hair zinc concentrations (24c.6)
8: Introduction to Anthropometry (Chapter 9)
8.1: Measurements, indices, and indicators
8.4: Interpretation and evaluation of anthropometric data
9: Body size (Chapter 10)
9.3: Growth indices, indicators, and recommended growth reference data (10.2)
mass spectrometry
12.4: Total body water from isotope dilution (14.3)
18.2: Biochemical indices of thiamine status (20a.2)
maternal health
14.9: Serum soluble transferrin receptor (17.9)
maternal nutrition
21: Folate (Chapter 22a)
MCV
14.4: Red-cell indices
14.5: Red-cell distribution width (17.5)
measurement
13: Biomarkers (Chapter 15)
24.3: Interpretive criterion- serum phosphate (23b.3)
measurement error
12.3: Total body potassium (TBK) (14.2)
4: Measurement Error in Dietary assessment (Chapter 5)
4.1: Measurement error in dietary intake data
4.3: Implications of measurement error for estimated dietary intakes
4.4: Minimizing measurement error through data collection procedures
4.5: Summary
8.2: Advantages and limitations of anthropometry
measurement errors
1.6: Important characteristics of assessment measures
14.3: Hematocrit or packed cell volume
8.3: Errors in anthropometry
measurement methods
12.7: Densitometry (14.6)
18.2: Biochemical indices of thiamine status (20a.2)
3: Food consumption of individuals
measurement techniques
12.4: Total body water from isotope dilution (14.3)
measurements
9.1: Anthropometric assessment of body size (10.0)
measurementtechniques
12.9: Bioelectrical impedance (14.8)
meats
25.5: Food sources and dietary Intakes
medical conditions
27.1: Copper (24b.0)
medical imaging
12.11: Magnetic resonance imaging (14.10)
MediterraneanDiet
7.5: Mediterranean Diet Score or Index (8c.5)
megaloblastic anemia
22.1: Introduction (22b.1)
megoblastic anemia
21: Folate (Chapter 22a)
memory errors
4.2: Sources of measurement error
Menkes syndrome
27.2: Serum copper (24b.1)
metaAnalysis
16.7: Interpretive criteria (18b.8)
metabolic defect
13.3: Biomarkers of status (15.3)
metabolic diseases
20: Niacin (Chapter 20c)
metabolic disorders
13.4: Biomarkers of function (15.4)
metabolic disturbances
19.2: Biomarkers of riboflavin status (20b.2)
metabolic risk
10.2: Assessment of body fat (11.1)
metabolic studies
25.8: Indices of magnesium status (23c.8)
metabolic syndrome
8.1: Measurements, indices, and indicators
9.4: Body mass index in adults (10.3)
metabolism
14.1: Introduction and functions of iron (17.1)
18.1: Thiamine (20a.1)
19: Riboflavin (Chapter 20b)
19.1: Riboflavin (20b.1)
20.1: Niacin - Introduction (20c.1)
20.2: Indices of niacin status (20c.2)
22.1: Introduction (22b.1)
27: Copper (Chapter 24b)
27.1: Copper (24b.0)
27.2: Serum copper (24b.1)
27.3: Serum ceruloplasmin (24b.2)
27.9: Other indices of copper status (24b.8)
28: Zinc (Chapter 24c)
28.1: Zinc (24c.1)
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
29.2: Selenium Chemistry (25b.2)
29.7: Metabolism (25b.7)
29.9: Assessing Selenium Status (25b.9)
metabolites
13: Biomarkers (Chapter 15)
26.3: Absorption and metabolism (23c.3)
metabolomics
13.7: Application of new technologies (15.7)
metalloenzymes
17.2: Functions of Vitamin C (19.2)
28.14: Zinc-dependent enzymes (24c.14)
metallothionein
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)

methionine synthesis
21: Folate (Chapter 22a)

methyl group donor
21: Folate (Chapter 22a)

methylation reactions
21.3: Conclusions (22a.3)

methylmalonic acid
22.2: Biomarkers of vitamin B12 status (22b.2)

microbiome
29.3: Health Effects (25b.3)

microcephalic
9.2: Measurements of body size (10.1)

microcytic
14: Iron (Chapter 17)

micronutrient
1.4: Nutritional assessment indices and indicators
14.1: Introduction and functions of iron (17.1)
7.10: Lower-burden diet quality indices (8c.10)

micronutrient tests
1.1: New developments in nutritional assessment

Micronutrients
13.5: Factors affecting choice of nutritional biomarkers (15.5)
5: Nutrient reference values (Chapter 8a)
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

microorganisms
24.1: Phosphorus (23b.1)

milk
23.4: Dietary sources and intakes of calcium (23a.4)

mineral metabolism
24: Phosphorus (Chapter 23b)

minerals
23.1: Calcium (23a.1)
5.4: U.K. Dietary Reference Values (8a.4)
5.6: European Dietary Reference Values (8a.6)
5.7: WHO/FAO Nutrient Reference Values (8a.7)

missense mutations
24.9: Inherited and tumor-induced disorders of phosphate metabolism

mitochondrial_activity
27.5: Cytochrome c oxidase (24b.4)

mixed diets
25.5: Food sources and dietary Intakes

mobile apps
3.2: Technical improvements in food consumption measurements

monocytes
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
28.16: Leukocyte zinc (24c.16)

MRI
12.11: Magnetic resonance imaging (14.10)

mRNA
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)

multiple dilution
12.5: Multiple dilution methods (14.4)

Muscle Contraction
26: Magnesium (Chapter 23c)

muscle mass
10.1: Anthropometric assessment of body composition (11.0)
10.3: Assessment of fat-free mass (11.2)
12.3: Total body potassium (TBK) (14.2)

muscle measurement
12.13: Ultrasound (14.12)

muscle spasms
25.4: Magnesium deficiency in humans
26.4: Magnesium deficiency in humans (23c.4)

muscle wasting
12.1: An introduction to techniques used to measure body composition (14.0)

mushrooms
16.4: Food sources and dietary intakes (18b.4)

myosteatosis
12.10: Computerized tomography (14.9)

N

NAA
12.6: In vivo activation analysis (14.5)

NAD
20: Niacin (Chapter 20c)
20.1: Niacin - Introduction (20c.1)
20.2: Indices of niacin status (20c.2)

nail zinc
28.7: Nail zinc (24c.7)

nausea
26.6: Effects of high magnesium intakes (23c.6)

NCD
7.10: Lower-burden diet quality indices (8c.10)

NCD prevention
7.1: Food-based dietary guidelines (8c.1)

NCD risk
7.12: Summary of diet quality indices (8c.12)

NCDs
7.4: Alternate Healthy Eating Index (8c.4)
7.6: Dietary Approaches to Stop Hypertension scores
7.8: Dietary Inflammatory Indices (8c.8)

nerve transmission
26: Magnesium (Chapter 23c)

neurobehavioral function
28: Zinc (Chapter 24c)

neurological
22.1: Introduction (22b.1)

neurological functioning
18: Thiamine (Chapter 20a)

neurological_disease
27.5: Cytochrome c oxidase (24b.4)

neuropsychiatric
25.4: Magnesium deficiency in humans

neurotransmitter
17.2: Functions of Vitamin C (19.2)

neutrophils
28.16: Leukocyte zinc (24c.16)

newborn growth
11.1: Introduction (13.0)

niacin
20: Niacin (Chapter 20c)
20.1: Niacin - Introduction (20c.1)
20.2: Indices of niacin status (20c.2)

niacin deficiency
20.2: Indices of niacin status (20c.2)

nicotinamide
20: Niacin (Chapter 20c)

nicotinic acid
20.1: Niacin - Introduction (20c.1)

noncommunicable diseases
24.6: Nutrient reference values (23b.6)

noninvasive
23.9: Measurement of bone mineral content and bone mineral density (23a.9)

noninvasive method
12.13: Ultrasound (14.12)

noninvasive methods
28.17: Salivary zinc (24c.17)

normocalcemic
16.11: Serum parathyroid hormone (18b.13)

NOVA classification
7.9: Diet quality based on level of food processing (8c.9)

NRVs
5: Nutrient reference values (Chapter 8a)
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)

nutrient
5: Nutrient reference values (Chapter 8a)
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

nutrient absorption
24.4: Phosphate balance (23b.4)

nutrient adequacy
28.2: Dietary Zinc Intake (24c.2)
5.5: United States and Canada- Dietary Reference Intakes (8a.5)
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)
7.12: Summary of diet quality indices (8c.12)
7.2: Assessing diet quality (8c.2)

nutrient assessment
28.14: Zinc-dependent enzymes (24c.14)

nutrient composition
4.1: Measurement error in dietary intake data

nutrient deficiencies
1.3: Nutritional assessment methods

nutrient deficiency
13.4: Biomarkers of function (15.4)
8.2: Advantages and limitations of anthropometry

nutrient deficits
1.1: New developments in nutritional assessment

nutrient intake
13.2: Biomarkers of exposure (15.2)
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
3.1: Methods for measuring food consumption of individuals
3.3: Selecting an appropriate method
4.5: Summary
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
6: Evaluation of intakes and diets (Chapter 8b)
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

nutrient intake assessment
5.5: United States and Canada- Dietary Reference Intakes (8a.5)

nutrient intakes
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
6.4: Using the NRVs for planning nutritious diets (8b.4)

nutrient reference values
5.1: Development of Nutrient Reference Values (8a.1)
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
5.4: U.K. Dietary Reference Values (8a.4)
5.5: United States and Canada- Dietary Reference Intakes (8a.5)
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
6: Evaluation of intakes and diets (Chapter 8b)
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

nutrient requirement
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

nutrient status
13.1: Biomarkers to assess nutritional status (15.1)

nutrient supply
2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES)

nutrient_deficiency
28.4: Linear growth (24c.4)

nutrientdistribution
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

nutrientrequirements
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

nutrients
13.3: Biomarkers of status (15.3)
24.1: Phosphorus (23b.1)
5.6: European Dietary Reference Values (8a.6)
5.7: WHO/FAO Nutrient Reference Values (8a.7)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)

nutrientvariability
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

nutrigenetics
13.7: Application of new technologies (15.7)

nutrition
1.3: Nutritional assessment methods
1.7: Evaluation of nutritional assessment indices
13.4: Biomarkers of function (15.4)
13.5: Factors affecting choice of nutritional biomarkers (15.5)
14.1: Introduction and functions of iron (17.1)
14.2: Hemoglobin (17.2)
15.1: Introduction (18a.1)
15.2: Indices of vitamin A status (18a.2)
16.4: Food sources and dietary intakes (18b.4)
17.5: Food sources and dietary intakes of vitamin C (19.5)
17.7: Indices of vitamin C status (19.7)
18.1: Thiamine (20a.1)
2.2: Food Balance Sheets
21.2: Biomarkers of folate status (Chapter 22a.2)
23.4: Dietary sources and intakes of calcium (23a.4)
23.5: Nutrient reference values for calcium (23a.5)
23.7: Effects of high intakes of calcium (23a.7)
24: Phosphorus (Chapter 23b)
25.1: Introduction (23c.1)
26.7: Occurrence of dietary deficiency (23c.7)
27.1: Copper (24b.1)
27.4: Erythrocyte superoxide dismutase (24b.3)
28.16: Leukocyte zinc (24c.16)
28.17: Salivary zinc (24c.17)
28.18: Multiple biomarkers (24c.18)
29: Selenium (Chapter 25b)
29.1: Overview (25b.1)
29.10: Questions Remain (25b.10)
29.3: Health Effects (25b.3)
29.4: Selenium Requirements (25b.4)
29.5: Food Sources (25b.5)
29.6: Enteric Absorption (25b.6)
29.9: Assessing Selenium Status (25b.9)
3.1: Methods for measuring food consumption of individuals
3.2: Technical improvements in food consumption measurements
5.4: U.K. Dietary Reference Values (8a.4)
5.6: European Dietary Reference Values (8a.6)
6.4: Using the NRVs for planning nutritious diets (8b.4)
7: Dietary guidelines and quality (Chapter 8c)
7.1: Food-based dietary guidelines (8c.1)
7.10: Lower-burden diet quality indices (8c.10)
7.2: Assessing diet quality (8c.2)
7.4: Alternate Healthy Eating Index (8c.4)
7.5: Mediterranean Diet Score or Index (8c.5)
7.7: Healthy Diet Indicator (8c.7)
7.8: Dietary Inflammatory Indices (8c.8)
9.4: Body mass index in adults (10.3)

nutrition analysis
2.3: Household Surveys

nutrition assessment
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

nutrition data
4.2: Sources of measurement error

nutrition deficiency
25.1: Introduction (23c.1)

nutrition education
1.2: Nutritional assessment systems
3.3: Selecting an appropriate method
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

nutrition intervention
9.1: Anthropometric assessment of body size (10.0)

nutrition monitoring
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

nutrition policy
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)

nutrition research
4.4: Minimizing measurement error through data collection procedures

nutrition surveys
1.2: Nutritional assessment systems

nutritional
13: Biomarkers (Chapter 15)
16.3: Vitamin D deficiency in humans (18b.3)

nutritional adequacy
6.4: Using the NRVs for planning nutritious diets (8b.4)

nutritional anthropometry
8.3: Errors in anthropometry

nutritional assessment
1: Introduction
1.1: New developments in nutritional assessment
1.2: Nutritional assessment systems
1.4: Nutritional assessment indices and indicators
1.6: Important characteristics of assessment measures
10.3: Assessment of fat-free mass (11.2)
19.2: Biomarkers of riboflavin status (20b.2)
2.1: Introduction
22.2: Biomarkers of vitamin B12 status (22b.2)
4.3: Implications of measurement error for estimated dietary intakes
7.9: Diet quality based on level of food processing (8c.9)
8.1: Measurements, indices, and indicators
8.2: Advantages and limitations of anthropometry
9.5: BMI in children and adolescents (10.4)

nutritional biomarkers
13.1: Biomarkers to assess nutritional status (15.1)
13.6: Evaluation of the selected nutritional biomarker (15.6)

nutritional deficiencies
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)

nutritional deficiency
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)

nutritional guidelines
28.2: Dietary Zinc Intake (24c.2)

nutritional research
13.2: Biomarkers of exposure (15.2)
26.8: Indices of magnesium status (23c.8)

nutritional science
24.2: Biological forms of phosphate and their measurement (23b.2)

nutritional standards
5.7: WHO/FAO Nutrient Reference Values (8a.7)

nutritional status
1.2: Nutritional assessment systems
11.1: Introduction (13.0)
11.2: Anthropometric indices and their modes of expression (13.1)
12.5: Multiple dilution methods (14.4)
13.3: Biomarkers of status (15.3)
13.5: Factors affecting choice of nutritional biomarkers (15.5)
20.2: Indices of niacin status (20c.2)
26: Magnesium (Chapter 23c)
8: Introduction to Anthropometry (Chapter 9)
8.4: Interpretation and evaluation of anthropometric data
9.3: Growth indices, indicators, and recommended growth reference data (10.2)

nutritional support
10.1: Anthropometric assessment of body composition (11.0)
12.6: In vivo activation analysis (14.5)

nutritional surveillance
8: Introduction to Anthropometry (Chapter 9)
8.4: Interpretation and evaluation of anthropometric data

nutritional_status
28.4: Linear growth (24c.4)

nutritionalassessment
1.5: The design of nutritional assessment systems

nutritionalFactors
17: Vitamin C (Chapater 19)

nutritionalguidelines
7.3: Examples of diet quality indices (8c.3)

nutritionalintake
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

nutritionresearch
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)

nuts
26.5: Food sources and dietary Intakes (23c.5)
7.4: Alternate Healthy Eating Index (8c.4)

O

obesity
1.4: Nutritional assessment indices and indicators
1.7: Evaluation of nutritional assessment indices
10: Body composition (Chapter 11)
10.1: Anthropometric assessment of body composition (11.0)
10.2: Assessment of body fat (11.1)
10.3: Assessment of fat-free mass (11.2)
11: Evaluation of Anthropometric Data (Chapter 13)
11.4: Screening systems to identify individuals at risk in a population (13.3)
12.1: An introduction to techniques used to measure body composition (14.0)
12.10: Computerized tomography (14.9)
12.12: Dual energy X-ray absorptiometry (14.11)
12.4: Total body water from isotope dilution (14.3)
12.7: Densitometry (14.6)
14.7: Serum ferritin (17.7)
17.8: Plasma ascorbic acid (19.8)
7.9: Diet quality based on level of food processing (8c.9)
8.1: Measurements, indices, and indicators
8.3: Errors in anthropometry
9.1: Anthropometric assessment of body size (10.0)
9.4: Body mass index in adults (10.3)
9.5: BMI in children and adolescents (10.4)

observational studies
3.1: Methods for measuring food consumption of individuals

older women
23.7: Effects of high intakes of calcium (23a.7)

OliveOil
7.5: Mediterranean Diet Score or Index (8c.5)

onychophagia
28.7: Nail zinc (24c.7)

organic compounds
24.2: Biological forms of phosphate and their measurement (23b.2)

organic phosphate
24.7: Dietary sources

osteoarthritis
25.4: Magnesium deficiency in humans

osteoblasts
23.2: Functions of calcium (23a.2)

osteoclasts
23.2: Functions of calcium (23a.2)

osteomalacia
16.10: Other vitamin D metabolites (18b.11)
16.12: Calcium and phosphorus in serum and urine (18b.14)
16.3: Vitamin D deficiency in humans (18b.3)
16.6: Biomarkers of vitamin D status (18b.6)
24.9: Inherited and tumor-induced disorders of phosphate metabolism

osteoporosis
12.12: Dual energy X-ray absorptiometry (14.11)
12.14: Summary - Method comparisons (14.13)
16: Vitamin D (Chapter 18b)
16.10: Other vitamin D metabolites (18b.11)
16.7: Interpretive criteria (18b.8)
23: Calcium (Chapter 23a)
23.6: Calcium deficiency (23a.6)
23.8: Biochemical indices of calcium status (23a.8)
23.9: Measurement of bone mineral content and bone mineral density (23a.9)
25.4: Magnesium deficiency in humans
25.8: Indices of magnesium status (23c.8)
26.4: Magnesium deficiency in humans (23c.4)

overload
27: Copper (Chapter 24b)

overnutrition
8.2: Advantages and limitations of anthropometry

overreporting
4.3: Implications of measurement error for estimated dietary intakes

overweight
11.2: Anthropometric indices and their modes of expression (13.1)
9.3: Growth indices, indicators, and recommended growth reference data (10.2)
9.4: Body mass index in adults (10.3)
9.5: BMI in children and adolescents (10.4)

oxidation
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
20: Niacin (Chapter 20c)

oxidative
27.3: Serum ceruloplasmin (24b.2)

oxidative damage
14.7: Serum ferritin (17.7)

oxidative stress
27.4: Erythrocyte superoxide dismutase (24b.3)
28: Zinc (Chapter 24c)
28.11: Oxidative stress and DNA integrity (24c.11)
29.9: Assessing Selenium Status (25b.9)

oxidative_stress
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)

oxygen
14.2: Hemoglobin (17.2)

P

pancreas
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)

parathyroid
23: Calcium (Chapter 23a)

parathyroid hormone
23.8: Biochemical indices of calcium status (23a.8)

parathyroid_hormone
16.11: Serum parathyroid hormone (18b.13)

parathyroidHormone
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)

pediatric health
11.3: Use of anthropometric indices in clinical settings (13.2)

pediatric nutrition
28: Zinc (Chapter 24c)

pellagra
20: Niacin (Chapter 20c)
percentage body fat
12.7: Densitometry (14.6)
percentiles
9: Body size (Chapter 10)
personalized nutrition
13.7: Application of new technologies (15.7)
pharmacokinetics
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
17.8: Plasma ascorbic acid (19.8)
phosphate
24: Phosphorus (Chapter 23b)
24.1: Phosphorus (23b.1)
24.2: Biological forms of phosphate and their measurement (23b.2)
24.3: Interpretive criterion- serum phosphate (23b.3)
24.4: Phosphate balance (23b.4)
24.7: Dietary sources
24.8: Abnormalities in serum phosphate (23b.8)
phosphate metabolism
24.9: Inherited and tumor-induced disorders of phosphate metabolism
phosphate wasting
24.9: Inherited and tumor-induced disorders of phosphate metabolism
phosphorus
16.1: Functions of vitamin D (18b.1)
16.12: Calcium and phosphorus in serum and urine (18b.14)
16.6: Biomarkers of vitamin D status (18b.6)
23.1: Calcium (23a.1)
24: Phosphorus (Chapter 23b)
24.1: Phosphorus (23b.1)
24.2: Biological forms of phosphate and their measurement (23b.2)
24.6: Nutrient reference values (23b.6)
25.3: Absorption and metabolism (23c.3)
physical performance
10.3: Assessment of fat-free mass (11.2)
physiological activity
25: Phosphorus (Chapter 23b)
physiological needs
14.2: Hemoglobin (17.2)
physiological samples
12.4: Total body water from isotope dilution (14.3)
physiological symptoms
28.14: Zinc-dependent enzymes (24c.14)
physiology
16.11: Serum parathyroid hormone (18b.13)
phytate
24.7: Dietary sources
28.2: Dietary Zinc Intake (24c.2)
pilot study
4.4: Minimizing measurement error through data collection procedures
planetary health
7.12: Summary of diet quality indices (8c.12)
planning diets
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
plant
25.5: Food sources and dietary Intakes
plant foods
26.5: Food sources and dietary Intakes (23c.5)
plant protein
24.7: Dietary sources
PlantBasedFoods
7.5: Mediterranean Diet Score or Index (8c.5)
plantfoods
29.5: Food Sources (25b.5)
plants
17.5: Food sources and dietary intakes of vitamin C (19.5)
plasma
17.12: Ascorbic acid in other cells and fluids (19.12)
17.7: Indices of vitamin C status (19.7)
26.8: Indices of magnesium status (23c.8)
28.15: Erythrocyte zinc (24c.15)
28.3: Serum zinc (24c.3)
plasma concentrations
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
plasma selenium levels
29.10: Questions Remain (25b.10)
plasma transport
24.4: Phosphate balance (23b.4)
plasma zinc
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
plasmaConcentrations
17: Vitamin C (Chapater 19)
platelets
27.5: Cytochrome c oxidase (24b.4)
plethysmography
12.7: Densitometry (14.6)
population
28.3: Serum zinc (24c.3)
population health
1.2: Nutritional assessment systems
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
5.1: Development of Nutrient Reference Values (8a.1)
8.4: Interpretation and evaluation of anthropometric data
population level
6: Evaluation of intakes and diets (Chapter 8b)
population studies
15.2: Indices of vitamin A status (18a.2)
18.2: Biochemical indices of thiamine status (20a.2)
22.2: Biomarkers of vitamin B12 status (22b.2)
population subgroup
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
populationGroups
16.5: Nutrient reference values for vitamin D (18b.5)
populationhealth
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)
portion size
4.5: Summary
postmenopausal
16.11: Serum parathyroid hormone (18b.13)
potassium
12.2: Chemical analysis of cadavers (14.1)
12.3: Total body potassium (TBK) (14.2)
poultry
7.4: Alternate Healthy Eating Index (8c.4)
poverty measurement
2.3: Household Surveys
precision
1.6: Important characteristics of assessment measures
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
predictive value
1.6: Important characteristics of assessment measures
pregnancy
21.1: Folate (22a.1)
23.3: Absorption and metabolism of calcium (23a.3)
23.5: Nutrient reference values for calcium (23a.5)
24.6: Nutrient reference values (23b.6)
27.3: Serum ceruloplasmin (24b.2)
28.5: Urine zinc (24c.5)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)
pregnantwomen
22: Vitamin B12 (Chapter 22b)
prenatal examination
12.13: Ultrasound (14.12)
prevalence of undernourishment
2.1: Introduction
prevalence study
14.10: Multiple biomarkers (17.10)
preventive zinc supplement
28.7: Nail zinc (24c.7)
processed foods
24.7: Dietary sources
processing level
7.9: Diet quality based on level of food processing (8c.9)
proliferation
16.1: Functions of vitamin D (18b.1)
protein
12.2: Chemical analysis of cadavers (14.1)
27.3: Serum ceruloplasmin (24b.2)
27.8: Copper chaperones (24b.7)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)
5.6: European Dietary Reference Values (8a.6)
proteomics
13.7: Application of new technologies (15.7)
protoporphyrin
14.8: Erythrocyte protoporphyrin
PTH
16.11: Serum parathyroid hormone (18b.13)
24.5: Hormonal regulation of serum phosphate (23b.5)

public health
1: Introduction
1.2: Nutritional assessment systems
1.3: Nutritional assessment methods
1.4: Nutritional assessment indices and indicators
11: Evaluation of Anthropometric Data (Chapter 13)
11.1: Introduction (13.1)
11.4: Screening systems to identify individuals at risk in a population (13.3)
11.5: Applications of anthropometry in public health (13.4)
13.6: Evaluation of the selected nutritional biomarker (15.6)
14.1: Introduction and functions of iron (17.1)
14.2: Hemoglobin (17.2)
15.1: Introduction (18a.1)
15.2: Indices of vitamin A status (18a.2)
18: Thiamine (Chapter 20a)
2.5: National and household surveys in the global assessment of food security and diet quality by age and sex
25: Phosphorus (Chapter 23b)
26: Magnesium (Chapter 23c)
29.3: Health Effects (25b.3)
5.6: European Dietary Reference Values (8a.6)
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)
7.1: Food-based dietary guidelines (8c.1)
7.2: Assessing diet quality (8c.2)
7.7: Healthy Diet Indicator (8c.7)
8.1: Measurements, indices, and indicators
8.4: Interpretation and evaluation of anthropometric data
9: Body size (Chapter 10)
9.3: Growth indices, indicators, and recommended growth reference data (10.2)

public health nutrition
5.1: Development of Nutrient Reference Values (8a.1)

public_health
28.4: Linear growth (24c.4)

publichealth
16.7: Interpretive criteria (18b.8)
18.3: Conclusions (20a.3)
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)
7.3: Examples of diet quality indices (8c.3)

publichealthnutrition
7.10: Lower-burden diet quality indices (8c.10)

pulses
26.5: Food sources and dietary Intakes (23c.5)

putrescine
27.7: Diamine oxidase (24b.6)

Q

QCT
23.9: Measurement of bone mineral content and bone mineral density (23a.9)

quality
2: Food available for consumption at National and Household Levels

quality control
13.5: Factors affecting choice of nutritional biomarkers (15.5)
28.6: Hair zinc concentrations (24c.6)
4.4: Minimizing measurement error through data collection procedures

quantitative assessment
2.1: Introduction

quantitative computed tomography
23.9: Measurement of bone mineral content and bone mineral density (23a.9)

R

radiation exposure
12.10: Computerized tomography (14.9)
12.6: In vivo activation analysis (14.5)

radioactive isotopes
17.13: Body pool size (19.13)

random error
4: Measurement Error in Dietary assessment (Chapter 5)
4.4: Minimizing measurement error through data collection procedures

randomized controlled trial
7.6: Dietary Approaches to Stop Hypertension scores

RBC folate
21.3: Conclusions (22a.3)

RDA
23.5: Nutrient reference values for calcium (23a.5)
26.7: Occurrence of dietary deficiency (23c.7)
29.4: Selenium Requirements (25b.4)

RDAs
17.6: Nutrient reference values for vitamin C (19.6)

RDW
14.5: Red-cell distribution width (17.5)

reabsorption
25.3: Absorption and metabolism (23c.3)

reactive oxygen species
28.11: Oxidative stress and DNA integrity (24c.11)

recall bias
4.2: Sources of measurement error

recognition thresholds
28.13: Taste acuity (24c.13)

Recommendations
25.7: Occurrence of dietary deficiency (23c.7)

recommended dietary allowance
24.6: Nutrient reference values (23b.6)
26.7: Occurrence of dietary deficiency (23c.7)

Recommended Dietary Allowances
5.1: Development of Nutrient Reference Values (8a.1)

recommended intake
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
5.7: WHO/FAO Nutrient Reference Values (8a.7)
6: Evaluation of intakes and diets (Chapter 8b)

recommendedIntakes
17.6: Nutrient reference values for vitamin C (19.6)

recovery biomarkers
13.2: Biomarkers of exposure (15.2)

red blood cells
14.3: Hematocrit or packed cell volume

redcellindices
14.4: Red-cell indices

redox
29.2: Selenium Chemistry (25b.2)

redox signalling
29.8: Biochemical Functions (25b.8)

redox state
28.11: Oxidative stress and DNA integrity (24c.11)

reduction
20: Niacin (Chapter 20c)

reference data
8: Introduction to Anthropometry (Chapter 9)

reference nutrient intake
5.4: U.K. Dietary Reference Values (8a.4)

reference values
1.7: Evaluation of nutritional assessment indices
5: Nutrient reference values (Chapter 8a)

regulation
24.5: Hormonal regulation of serum phosphate (23b.5)

renal excretion
23.3: Absorption and metabolism of calcium (23a.3)
24.4: Phosphate balance (23b.4)
24.8: Abnormalities in serum phosphate (23b.8)

renal failure
16.3: Vitamin D deficiency in humans (18b.3)

renal function
24: Phosphorus (Chapter 23b)

renal stones
23.7: Effects of high intakes of calcium (23a.7)

renal threshold
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)

renalaction
17.3: Absorption and metabolism of vitamin C (19.3)

reproducibility
1.6: Important characteristics of assessment measures
4.1: Measurement error in dietary intake data
4.4: Minimizing measurement error through data collection procedures

reproductive age
5: Nutrient reference values (Chapter 8a)

research
18.3: Conclusions (20a.3)
28.17: Salivary zinc (24c.17)
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)
7.3: Examples of diet quality indices (8c.3)

research methodology
8.3: Errors in anthropometry

researchguidelines
1.5: The design of nutritional assessment systems

respiratoryinfections
17.4: Deficiency of vitamin C in humans (19.4)

retinoids
15.1: Introduction (18a.1)

riboflavin
19: Riboflavin (Chapter 20b)
19.1: Riboflavin (20b.1)
19.2: Biomarkers of riboflavin status (20b.2)

rickets
16.10: Other vitamin D metabolites (18b.11)
16.3: Vitamin D deficiency in humans (18b.3)
16.6: Biomarkers of vitamin D status (18b.6)
24.9: Inherited and tumor-induced disorders of phosphate metabolism

risk
28.3: Serum zinc (24c.3)

risk assessment
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)
RNA
26.2: Functions of magnesium (23c.2)
RNA synthesis
25.2: Functions of magnesium (23c.2)

S

safe upper level
5.1: Development of Nutrient Reference Values (8a.1)
safe upper levels
5.4: U.K. Dietary Reference Values (8a.4)
saliva
17.12: Ascorbic acid in other cells and fluids (19.12)
28.17: Salivary zinc (24c.17)
sampling
1.5: The design of nutritional assessment systems
sarcopenia
10.3: Assessment of fat-free mass (11.2)
12.12: Dual energy X-ray absorptiometry (14.11)
12.14: Summary - Method comparisons (14.13)
scanning techniques
12: Body Composition- Laboratory Methods (Chapter 14)
12.1: An introduction to techniques used to measure body composition (14.0)
screening
13.6: Evaluation of the selected nutritional biomarker (15.6)
screening interventions
8: Introduction to Anthropometry (Chapter 9)
screening systems
11.1: Introduction (13.0)
screening tools
11.4: Screening systems to identify individuals at risk in a population (13.3)
Scurvy
17: Vitamin C (Chapater 19)
17.1: Vitamin C (19.1)
17.14: Functional tests of vitamin C status (19.14)
17.2: Functions of Vitamin C (19.2)
17.4: Deficiency of vitamin C in humans (19.4)
Se deficiency
29.3: Health Effects (25b.3)
Se intake
29.9: Assessing Selenium Status (25b.9)
seasonalVariation
17.5: Food sources and dietary intakes of vitamin C (19.5)
security
2: Food available for consumption at National and Household Levels
selenate
29.2: Selenium Chemistry (25b.2)
29.6: Enteric Absorption (25b.6)
selenide
29.2: Selenium Chemistry (25b.2)
selenite
29.2: Selenium Chemistry (25b.2)
29.6: Enteric Absorption (25b.6)
29.7: Metabolism (25b.7)
selenium
29: Selenium (Chapter 25b)
29.1: Overview (25b.1)
29.10: Questions Remain (25b.10)
29.2: Selenium Chemistry (25b.2)
29.3: Health Effects (25b.3)
29.4: Selenium Requirements (25b.4)
29.5: Food Sources (25b.5)
29.6: Enteric Absorption (25b.6)
29.7: Metabolism (25b.7)
29.9: Assessing Selenium Status (25b.9)
selenium deficiency
29.5: Food Sources (25b.5)
selenium exposure
28.7: Nail zinc (24c.7)
selenocysteine
29: Selenium (Chapter 25b)
29.2: Selenium Chemistry (25b.2)
29.7: Metabolism (25b.7)
selenocysteine
29.8: Biochemical Functions (25b.8)
selenomethionine
29.2: Selenium Chemistry (25b.2)
29.6: Enteric Absorption (25b.6)
29.7: Metabolism (25b.7)
selenoproteins
29: Selenium (Chapter 25b)
29.7: Metabolism (25b.7)
29.8: Biochemical Functions (25b.8)
selenosis
29.7: Metabolism (25b.7)
sensitivity
1.6: Important characteristics of assessment measures
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
serine metabolism
21: Folate (Chapter 22a)
serum
16.3: Vitamin D deficiency in humans (18b.3)
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
24.3: Interpretive criterion- serum phosphate (23b.3)
26.8: Indices of magnesium status (23c.8)
27: Copper (Chapter 24b)
27.2: Serum copper (24b.1)
27.7: Diamine oxidase (24b.6)
27.9: Other indices of copper status (24b.8)
28.3: Serum zinc (24c.3)
serum calcium
23.6: Calcium deficiency (23a.6)
serum concentration
24.4: Phosphate balance (23b.4)
serum copper
27.4: Erythrocyte superoxide dismutase (24b.3)
serum ferritin
14.10: Multiple biomarkers (17.10)
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
serum folate
21.3: Conclusions (22a.3)
serum magnesium
25: Phosphorus (Chapter 23b)
serum phosphate
24.8: Abnormalities in serum phosphate (23b.8)
serum retinol
15.2: Indices of vitamin A status (18a.2)
serum_levels
16.11: Serum parathyroid hormone (18b.13)
serum_phosphate
24.5: Hormonal regulation of serum phosphate (23b.5)
serum25hydroxyvitaminD
16.6: Biomarkers of vitamin D status (18b.6)
serum25OHD
16.5: Nutrient reference values for vitamin D (18b.5)
16.7: Interpretive criteria (18b.8)
serumConcentration
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)
sheets
2: Food available for consumption at National and Household Levels
sickle cell disease
14.5: Red-cell distribution width (17.5)
skeletal
25.1: Introduction (23c.1)
26.1: Introduction (23c.1)
skeletal abnormalities
27.6: Skin lysyl oxidase (24b.5)
skeletal health
23.9: Measurement of bone mineral content and bone mineral density (23a.9)
Skeletal Muscle
12.10: Computerized tomography (14.9)
12.11: Magnetic resonance imaging (14.10)
skeletal system
23.1: Calcium (23a.1)
skeletalhealth
16.6: Biomarkers of vitamin D status (18b.6)
skinfolds
10: Body composition (Chapter 11)
smoking
17.8: Plasma ascorbic acid (19.8)
SNP
13.7: Application of new technologies (15.7)
socioeconomic status
4.3: Implications of measurement error for estimated dietary intakes
SOD1
27.8: Copper chaperones (24b.7)
sodiumdependenttransporter
17.3: Absorption and metabolism of vitamin C (19.3)
soil
24.1: Phosphorus (23b.1)
soluble transferrin receptor
14.10: Multiple biomarkers (17.10)
sonography
12.13: Ultrasound (14.12)
special needs
6.4: Using the NRVs for planning nutritious diets (8b.4)
Specificity
1.6: Important characteristics of assessment measures
standardization
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
18.3: Conclusions (20a.3)
statistical methods
6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2)

statistical techniques
3.3: Selecting an appropriate method
statisticalanalysis
1.5: The design of nutritional assessment systems
status
16.12: Calcium and phosphorus in serum and urine (18b.14)
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
storage
14.1: Introduction and functions of iron (17.1)
stress
27: Copper (Chapter 24b)
study design
4: Measurement Error in Dietary assessment (Chapter 5)
study objectives
3.3: Selecting an appropriate method
stunting
1.4: Nutritional assessment indices and indicators
11: Evaluation of Anthropometric Data (Chapter 13)
11.2: Anthropometric indices and their modes of expression (13.1)
11.5: Applications of anthropometry in public health (13.4)
28.18: Multiple biomarkers (24c.18)
28.4: Linear growth (24c.4)
9: Body size (Chapter 10)
9.1: Anthropometric assessment of body size (10.0)
9.3: Growth indices, indicators, and recommended growth reference data (10.2)
sunlightExposure
16.5: Nutrient reference values for vitamin D (18b.5)
superoxide dismutase
27.4: Erythrocyte superoxide dismutase (24b.3)
supplementation
16.6: Biomarkers of vitamin D status (18b.6)
17.14: Functional tests of vitamin C status (19.14)
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
20.1: Niacin - Introduction (20c.1)
25.8: Indices of magnesium status (23c.8)
27.4: Erythrocyte superoxide dismutase (24b.3)
27.6: Skin lysyl oxidase (24b.5)
27.7: Diamine oxidase (24b.6)
28.1: Zinc (24c.1)
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
28.3: Serum zinc (24c.3)
28.5: Urine zinc (24c.5)
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)
29: Selenium (Chapter 25b)
29.1: Overview (25b.1)
29.9: Assessing Selenium Status (25b.9)
supplementation trials
28.14: Zinc-dependent enzymes (24c.14)
supplements
16.4: Food sources and dietary intakes (18b.4)
22.1: Introduction (22b.1)
23.4: Dietary sources and intakes of calcium (23a.4)
23.7: Effects of high intakes of calcium (23a.7)
25.6: Effects of high magnesium intakes (23c.6)
29.4: Selenium Requirements (25b.4)
surveillance systems
1: Introduction
survey methods
2.1: Introduction
4.1: Measurement error in dietary intake data
surveys
3.2: Technical improvements in food consumption measurements
sustainability
7: Dietary guidelines and quality (Chapter 8c)
7.2: Assessing diet quality (8c.2)
sustainable diets
7.11: Incorporating sustainability in diet quality definitions and indices (8c.11)
symptoms
18.1: Thiamine (20a.1)
25.4: Magnesium deficiency in humans
25.6: Effects of high magnesium intakes (23c.6)
26.6: Effects of high magnesium intakes (23c.6)
systematic error
4: Measurement Error in Dietary assessment (Chapter 5)
4.4: Minimizing measurement error through data collection procedures
systematic review
5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2)

T

T2D risk
29.10: Questions Remain (25b.10)
taste acuity
28.13: Taste acuity (24c.13)
28.15: Erythrocyte zinc (24c.15)
28.17: Salivary zinc (24c.17)
taste qualities
28.13: Taste acuity (24c.13)
technical improvements
3: Food consumption of individuals
techniques
8.3: Errors in anthropometry
technological limitations
12.14: Summary - Method comparisons (14.13)
technology
13: Biomarkers (Chapter 15)
3.2: Technical improvements in food consumption measurements
4.4: Minimizing measurement error through data collection procedures
technology enabled methods
4.2: Sources of measurement error
teeth
23: Calcium (Chapter 23a)
23.2: Functions of calcium (23a.2)
thalassemia
14.11: Emerging iron status indicators (17.11)
14.5: Red-cell distribution width (17.5)
therapeutic thiamine challenge
18: Thiamine (Chapter 20a)
therapeutic zinc dose
28.7: Nail zinc (24c.7)
thiamine
18: Thiamine (Chapter 20a)
18.1: Thiamine (20a.1)
18.2: Biochemical indices of thiamine status (20a.2)
18.3: Conclusions (20a.3)
thiamine deficiency
18: Thiamine (Chapter 20a)
Thioredoxin Reductases
29.8: Biochemical Functions (25b.8)
TIBC
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
tissue thickness
12.13: Ultrasound (14.12)
tissues
13.3: Biomarkers of status (15.3)
tissueuptake
17.3: Absorption and metabolism of vitamin C (19.3)
TOBEC
12.8: Total body electrical conductivity (14.7)
total body calcium
12.6: In vivo activation analysis (14.5)
total body nitrogen
12.6: In vivo activation analysis (14.5)
total body water
12: Body Composition- Laboratory Methods (Chapter 14)
12.4: Total body water from isotope dilution (14.3)
12.5: Multiple dilution methods (14.4)
toxicity
25.6: Effects of high magnesium intakes (23c.6)
26.6: Effects of high magnesium intakes (23c.6)
29.1: Overview (25b.1)
trace elements
27.1: Copper (24b.0)
28.1: Zinc (24c.1)
28.6: Hair zinc concentrations (24c.6)
trade
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
training
8.3: Errors in anthropometry
transcription factor
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)
transferrin
14: Iron (Chapter 17)
14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6)
14.8: Erythrocyte protoporphyrin
transferrin receptor
14.9: Serum soluble transferrin receptor (17.9)
transferrin saturation
14.10: Multiple biomarkers (17.10)
transketolase
18.2: Biochemical indices of thiamine status (20a.2)
transparency
5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9)
transport
21.1: Folate (22a.1)
26.3: Absorption and metabolism (23c.3)
29.6: Enteric Absorption (25b.6)
transporters
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
treatment
13.6: Evaluation of the selected nutritional biomarker (15.6)
TRPM6
25.3: Absorption and metabolism (23c.3)

turnover rate
26.1: Introduction (23c.1)
turnover rates
25.1: Introduction (23c.1)
type 2 diabetes
7.6: Dietary Approaches to Stop Hypertension scores

U

U.S. Food and Nutrition Board
5.5: United States and Canada- Dietary Reference Intakes (8a.5)
UL
23.5: Nutrient reference values for calcium (23a.5)
ULs
17.6: Nutrient reference values for vitamin C (19.6)
ultrasound
12.1: An introduction to techniques used to measure body composition (14.0)
12.13: Ultrasound (14.12)
undernutrition
1.7: Evaluation of nutritional assessment indices
8.2: Advantages and limitations of anthropometry
underreporting
4.3: Implications of measurement error for estimated dietary intakes
underweight
11.2: Anthropometric indices and their modes of expression (13.1)
9.3: Growth indices, indicators, and recommended growth reference data (10.2)
upper intake level
6: Evaluation of intakes and diets (Chapter 8b)
urinary
27.9: Other indices of copper status (24b.8)
urinary analysis
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
urinary calcium excretion
23.3: Absorption and metabolism of calcium (23a.3)
urinary excretion
20.2: Indices of niacin status (20c.2)
25.8: Indices of magnesium status (23c.8)
26: Magnesium (Chapter 23c)
urinary metabolites
29.7: Metabolism (25b.7)
urinary zinc
28.5: Urine zinc (24c.5)
urinaryExcretion
17.14: Functional tests of vitamin C status (19.14)
urine
13.3: Biomarkers of status (15.3)
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
17.13: Body pool size (19.13)
urineexcretion
17.3: Absorption and metabolism of vitamin C (19.3)
UVB
16.2: Metabolism of vitamin D (18b.2)

V

validation studies
3.2: Technical improvements in food consumption measurements
validity
1.6: Important characteristics of assessment measures
4.1: Measurement error in dietary intake data
4.4: Minimizing measurement error through data collection procedures
8.3: Errors in anthropometry
vegetables
17.5: Food sources and dietary intakes of vitamin C (19.5)
25.5: Food sources and dietary Intakes
7.5: Mediterranean Diet Score or Index (8c.5)
7.6: Dietary Approaches to Stop Hypertension scores
7.7: Healthy Diet Indicator (8c.7)
venipuncture
14.3: Hematocrit or packed cell volume
visceral adipose tissue
10.1: Anthropometric assessment of body composition (11.0)
visceral fat
10: Body composition (Chapter 11)
10.2: Assessment of body fat (11.1)
9.4: Body mass index in adults (10.3)
vision
15.1: Introduction (18a.1)
vitamin A
15.1: Introduction (18a.1)
15.2: Indices of vitamin A status (18a.2)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)
vitamin A deficiency
15.2: Indices of vitamin A status (18a.2)
vitamin B1
18: Thiamine (Chapter 20a)
vitamin B12
13.4: Biomarkers of function (15.4)
13.6: Evaluation of the selected nutritional biomarker (15.6)
21: Folate (Chapter 22a)
21.2: Biomarkers of folate status (Chapter 22a.2)
22.1: Introduction (22b.1)
22.2: Biomarkers of vitamin B12 status (22b.2)
vitamin B12 deficiency
14.5: Red-cell distribution width (17.5)
21.3: Conclusions (22a.3)
vitamin B3
20.1: Niacin - Introduction (20c.1)
vitamin B6
13.4: Biomarkers of function (15.4)
vitamin C
17.1: Vitamin C (19.1)
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)
17.11: Urinary excretion of ascorbic acid and metabolites (19.11)
17.12: Ascorbic acid in other cells and fluids (19.12)
17.13: Body pool size (19.13)
17.8: Plasma ascorbic acid (19.8)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)
vitamin D
16: Vitamin D (Chapter 18b)
16.10: Other vitamin D metabolites (18b.11)
16.3: Vitamin D deficiency in humans (18b.3)
16.8: Measurement of serum 25-hydroxyvitamin D (18b.9)
23.2: Functions of calcium (23a.2)
23.3: Absorption and metabolism of calcium (23a.3)
23.6: Calcium deficiency (23a.6)
23.8: Biochemical indices of calcium status (23a.8)
26.3: Absorption and metabolism (23c.3)
5.8: Sources of discrepancies in Nutrient Reference Values (8a.8)
vitamin D deficiency
24.8: Abnormalities in serum phosphate (23b.8)
vitamin D2
16: Vitamin D (Chapter 18b)
vitamin D3
16: Vitamin D (Chapter 18b)
vitamin E
29.10: Questions Remain (25b.10)
vitamin_C
17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9)
vitamin_D
16.11: Serum parathyroid hormone (18b.13)
VitaminB12
14.4: Red-cell indices
22: Vitamin B12 (Chapter 22b)
vitaminB6
6.3: Using NRVs to assess nutrient intakes of groups (8b.3)
vitaminC
17: Vitamin C (Chapater 19)
17.14: Functional tests of vitamin C status (19.14)
17.2: Functions of Vitamin C (19.2)
17.3: Absorption and metabolism of vitamin C (19.3)
17.5: Food sources and dietary intakes of vitamin C (19.5)
17.6: Nutrient reference values for vitamin C (19.6)
17.7: Indices of vitamin C status (19.7)
vitaminCdeficiency
17.4: Deficiency of vitamin C in humans (19.4)
vitaminD
16.1: Functions of vitamin D (18b.1)
16.12: Calcium and phosphorus in serum and urine (18b.14)
16.2: Metabolism of vitamin D (18b.2)
16.4: Food sources and dietary intakes (18b.4)
16.5: Nutrient reference values for vitamin D (18b.5)
16.6: Biomarkers of vitamin D status (18b.6)
16.7: Interpretive criteria (18b.8)
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)
vitaminD2
16.2: Metabolism of vitamin D (18b.2)
vitaminD3
16.2: Metabolism of vitamin D (18b.2)
vitaminDDeficiency
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)

vitaminDSufficiency
16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10)

vitamins
14.9: Serum soluble transferrin receptor (17.9)
19.1: Riboflavin (20b.1)
5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3)
5.4: U.K. Dietary Reference Values (8a.4)
5.6: European Dietary Reference Values (8a.6)
5.7: WHO/FAO Nutrient Reference Values (8a.7)
6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1)

vitamins and minerals
5.5: United States and Canada- Dietary Reference Intakes (8a.5)

vitamins B12
21.1: Folate (22a.1)

vulnerable groups
1: Introduction

W

wasting
1.4: Nutritional assessment indices and indicators
11.2: Anthropometric indices and their modes of expression (13.1)
11.5: Applications of anthropometry in public health (13.4)
9: Body size (Chapter 10)
9.1: Anthropometric assessment of body size (10.0)
9.3: Growth indices, indicators, and recommended growth reference data (10.2)

water
12.2: Chemical analysis of cadavers (14.1)

weakness
17.4: Deficiency of vitamin C in humans (19.4)

weight
9: Body size (Chapter 10)
9.1: Anthropometric assessment of body size (10.0)

weight loss
12.1: An introduction to techniques used to measure body composition (14.0)

weight management
6.4: Using the NRVs for planning nutritious diets (8b.4)

weight status
4.3: Implications of measurement error for estimated dietary intakes

WHO
11.1: Introduction (13.0)
28.4: Linear growth (24c.4)
5.7: WHO/FAO Nutrient Reference Values (8a.7)
7.1: Food-based dietary guidelines (8c.1)
7.7: Healthy Diet Indicator (8c.7)

WHO Child Growth Standards
8.4: Interpretation and evaluation of anthropometric data

WHO guidelines
11.5: Applications of anthropometry in public health (13.4)

whole blood
17.10: Ascorbic acid in erythrocytes and whole blood (19.10)

whole grains
25.5: Food sources and dietary Intakes
26.5: Food sources and dietary Intakes (23c.5)
7.4: Alternate Healthy Eating Index (8c.4)
7.6: Dietary Approaches to Stop Hypertension scores

Wilson's disease
27.2: Serum copper (24b.1)
27.9: Other indices of copper status (24b.8)

Y

young healthy adults
17.13: Body pool size (19.13)

Z

zinc
13.2: Biomarkers of exposure (15.2)
13.4: Biomarkers of function (15.4)
27.8: Copper chaperones (24b.7)
28: Zinc (Chapter 24c)
28.1: Zinc (24c.1)
28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10)
28.11: Oxidative stress and DNA integrity (24c.11)
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
28.14: Zinc-dependent enzymes (24c.14)
28.15: Erythrocyte zinc (24c.15)
28.16: Leukocyte zinc (24c.16)
28.17: Salivary zinc (24c.17)
28.18: Multiple biomarkers (24c.18)
28.2: Dietary Zinc Intake (24c.2)
28.3: Serum zinc (24c.3)
28.4: Linear growth (24c.4)
28.5: Urine zinc (24c.5)
28.6: Hair zinc concentrations (24c.6)
28.8: Metallothionein concentrations in circulating blood cell types (24c.8)
28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9)
5: Nutrient reference values (Chapter 8a)

zinc absorption
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)

zinc concentrations
28.7: Nail zinc (24c.7)

zinc deficiency
28.13: Taste acuity (24c.13)
28.14: Zinc-dependent enzymes (24c.14)
28.2: Dietary Zinc Intake (24c.2)
28.6: Hair zinc concentrations (24c.6)

zinc depletion
28.15: Erythrocyte zinc (24c.15)

zinc exposure
28.7: Nail zinc (24c.7)

zinc status
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)
28.13: Taste acuity (24c.13)
28.14: Zinc-dependent enzymes (24c.14)
28.17: Salivary zinc (24c.17)

zinc supplementation
28.11: Oxidative stress and DNA integrity (24c.11)
28.12: Kinetic markers: pool sizes and turnover rates (24c.12)

This page titled Index is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson via source content that was edited to the style and standards of the LibreTexts platform.

DETAILED LICENSING

Overview

Title: Principles of Nutritional Assessment 3e (Gibson et al.)

Webpages: 244

All licenses found:

- CC BY 4.0: 99.2% (242 pages)
- Undeclared: 0.8% (2 pages)

By Page

- Principles of Nutritional Assessment 3e (Gibson et al.) - *CC BY 4.0*
  - Front Matter - *CC BY 4.0*
    - TitlePage - *CC BY 4.0*
    - InfoPage - *CC BY 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY 4.0*
  - 1: Introduction - *CC BY 4.0*
    - 1.1: New developments in nutritional assessment - *CC BY 4.0*
    - 1.2: Nutritional assessment systems - *CC BY 4.0*
    - 1.3: Nutritional assessment methods - *CC BY 4.0*
    - 1.4: Nutritional assessment indices and indicators - *CC BY 4.0*
    - 1.5: The design of nutritional assessment systems - *CC BY 4.0*
    - 1.6: Important characteristics of assessment measures - *CC BY 4.0*
    - 1.7: Evaluation of nutritional assessment indices - *CC BY 4.0*
  - 2: Food available for consumption at National and Household Levels - *CC BY 4.0*
    - 2.1: Introduction - *CC BY 4.0*
    - 2.2: Food Balance Sheets - *CC BY 4.0*
    - 2.3: Household Surveys - *CC BY 4.0*
    - 2.4: Comparison of Food Balance Sheets (FBS) and Household Consumption and Expenditure Surveys (HCES) - *CC BY 4.0*
    - 2.5: National and household surveys in the global assessment of food security and diet quality by age and sex - *CC BY 4.0*
  - 3: Food consumption of individuals - *CC BY 4.0*
    - 3.1: Methods for measuring food consumption of individuals - *CC BY 4.0*
    - 3.2: Technical improvements in food consumption measurements - *CC BY 4.0*
    - 3.3: Selecting an appropriate method - *CC BY 4.0*
  - 4: Measurement Error in Dietary assessment (Chapter 5) - *CC BY 4.0*
    - 4.1: Measurement error in dietary intake data - *CC BY 4.0*
    - 4.2: Sources of measurement error - *CC BY 4.0*
    - 4.3: Implications of measurement error for estimated dietary intakes - *CC BY 4.0*
    - 4.4: Minimizing measurement error through data collection procedures - *CC BY 4.0*
    - 4.5: Summary - *CC BY 4.0*
  - 5: Nutrient reference values (Chapter 8a) - *CC BY 4.0*
    - 5.1: Development of Nutrient Reference Values (8a.1) - *CC BY 4.0*
    - 5.2: Harmonization of methodological approaches to derive Nutrient Reference Values (8a.2) - *CC BY 4.0*
    - 5.3: A review of Nutrient Reference Values set by the UK, U.S./Canada, European Union, and WHO/FAO (8a.3) - *CC BY 4.0*
    - 5.4: U.K. Dietary Reference Values (8a.4) - *CC BY 4.0*
    - 5.5: United States and Canada- Dietary Reference Intakes (8a.5) - *CC BY 4.0*
    - 5.6: European Dietary Reference Values (8a.6) - *CC BY 4.0*
    - 5.7: WHO/FAO Nutrient Reference Values (8a.7) - *CC BY 4.0*
    - 5.8: Sources of discrepancies in Nutrient Reference Values (8a.8) - *CC BY 4.0*
    - 5.9: Future directions in harmonizing the process for deriving Nutrient Reference Values (8a.9) - *CC BY 4.0*
  - 6: Evaluation of intakes and diets (Chapter 8b) - *CC BY 4.0*
    - 6.1: Applications of Nutrient Reference Values in diet assessment and planning (8b.1) - *CC BY 4.0*
    - 6.2: Using the NRVs to assess nutrient intakes of an individual (8b.2) - *CC BY 4.0*
    - 6.3: Using NRVs to assess nutrient intakes of

groups (8b.3) - *CC BY 4.0*
- 6.4: Using the NRVs for planning nutritious diets (8b.4) - *CC BY 4.0*

- 7: Dietary guidelines and quality (Chapter 8c) - *CC BY 4.0*
  - 7.1: Food-based dietary guidelines (8c.1) - *CC BY 4.0*
  - 7.2: Assessing diet quality (8c.2) - *CC BY 4.0*
  - 7.3: Examples of diet quality indices (8c.3) - *CC BY 4.0*
  - 7.4: Alternate Healthy Eating Index (8c.4) - *CC BY 4.0*
  - 7.5: Mediterranean Diet Score or Index (8c.5) - *CC BY 4.0*
  - 7.6: Dietary Approaches to Stop Hypertension scores - *CC BY 4.0*
  - 7.7: Healthy Diet Indicator (8c.7) - *CC BY 4.0*
  - 7.8: Dietary Inflammatory Indices (8c.8) - *CC BY 4.0*
  - 7.9: Diet quality based on level of food processing (8c.9) - *CC BY 4.0*
  - 7.10: Lower-burden diet quality indices (8c.10) - *CC BY 4.0*
  - 7.11: Incorporating sustainability in diet quality definitions and indices (8c.11) - *CC BY 4.0*
  - 7.12: Summary of diet quality indices (8c.12) - *CC BY 4.0*
- 8: Introduction to Anthropometry (Chapter 9) - *CC BY 4.0*
  - 8.1: Measurements, indices, and indicators - *CC BY 4.0*
  - 8.2: Advantages and limitations of anthropometry - *CC BY 4.0*
  - 8.3: Errors in anthropometry - *CC BY 4.0*
  - 8.4: Interpretation and evaluation of anthropometric data - *CC BY 4.0*
- 9: Body size (Chapter 10) - *CC BY 4.0*
  - 9.1: Anthropometric assessment of body size (10.0) - *CC BY 4.0*
  - 9.2: Measurements of body size (10.1) - *CC BY 4.0*
  - 9.3: Growth indices, indicators, and recommended growth reference data (10.2) - *CC BY 4.0*
  - 9.4: Body mass index in adults (10.3) - *CC BY 4.0*
  - 9.5: BMI in children and adolescents (10.4) - *CC BY 4.0*
- 10: Body composition (Chapter 11) - *CC BY 4.0*
  - 10.1: Anthropometric assessment of body composition (11.0) - *CC BY 4.0*
  - 10.2: Assessment of body fat (11.1) - *CC BY 4.0*
  - 10.3: Assessment of fat-free mass (11.2) - *CC BY 4.0*
- 11: Evaluation of Anthropometric Data (Chapter 13) - *CC BY 4.0*
  - 11.1: Introduction (13.0) - *CC BY 4.0*
  - 11.2: Anthropometric indices and their modes of expression (13.1) - *CC BY 4.0*
  - 11.3: Use of anthropometric indices in clinical settings (13.2) - *CC BY 4.0*
  - 11.4: Screening systems to identify individuals at risk in a population (13.3) - *CC BY 4.0*
  - 11.5: Applications of anthropometry in public health (13.4) - *CC BY 4.0*
- 12: Body Composition- Laboratory Methods (Chapter 14) - *CC BY 4.0*
  - 12.1: An introduction to techniques used to measure body composition (14.0) - *CC BY 4.0*
  - 12.2: Chemical analysis of cadavers (14.1) - *CC BY 4.0*
  - 12.3: Total body potassium (TBK) (14.2) - *CC BY 4.0*
  - 12.4: Total body water from isotope dilution (14.3) - *CC BY 4.0*
  - 12.5: Multiple dilution methods (14.4) - *CC BY 4.0*
  - 12.6: In vivo activation analysis (14.5) - *CC BY 4.0*
  - 12.7: Densitometry (14.6) - *CC BY 4.0*
  - 12.8: Total body electrical conductivity (14.7) - *CC BY 4.0*
  - 12.9: Bioelectrical impedance (14.8) - *CC BY 4.0*
  - 12.10: Computerized tomography (14.9) - *CC BY 4.0*
  - 12.11: Magnetic resonance imaging (14.10) - *CC BY 4.0*
  - 12.12: Dual energy X-ray absorptiometry (14.11) - *CC BY 4.0*
  - 12.13: Ultrasound (14.12) - *CC BY 4.0*
  - 12.14: Summary - Method comparisons (14.13) - *CC BY 4.0*
- 13: Biomarkers (Chapter 15) - *CC BY 4.0*
  - 13.1: Biomarkers to assess nutritional status (15.1) - *CC BY 4.0*
  - 13.2: Biomarkers of exposure (15.2) - *CC BY 4.0*
  - 13.3: Biomarkers of status (15.3) - *CC BY 4.0*
  - 13.4: Biomarkers of function (15.4) - *CC BY 4.0*
  - 13.5: Factors affecting choice of nutritional biomarkers (15.5) - *CC BY 4.0*
  - 13.6: Evaluation of the selected nutritional biomarker (15.6) - *CC BY 4.0*
  - 13.7: Application of new technologies (15.7) - *CC BY 4.0*
- 14: Iron (Chapter 17) - *CC BY 4.0*
  - 14.1: Introduction and functions of iron (17.1) - *CC BY 4.0*
  - 14.2: Hemoglobin (17.2) - *CC BY 4.0*
  - 14.3: Hematocrit or packed cell volume - *CC BY 4.0*
  - 14.4: Red-cell indices - *CC BY 4.0*
  - 14.5: Red-cell distribution width (17.5) - *CC BY 4.0*
  - 14.6: Serum iron, TIBC, transferrin, and transferrin saturation (17.6) - *CC BY 4.0*

- 14.7: Serum ferritin (17.7) - *CC BY 4.0*
- 14.8: Erythrocyte protoporphyrin - *CC BY 4.0*
- 14.9: Serum soluble transferrin receptor (17.9) - *CC BY 4.0*
- 14.10: Multiple biomarkers (17.10) - *CC BY 4.0*
- 14.11: Emerging iron status indicators (17.11) - *CC BY 4.0*

◦ 15: Vitamin A (Chapter 18a) - *CC BY 4.0*
- 15.1: Introduction (18a.1) - *CC BY 4.0*
- 15.2: Indices of vitamin A status (18a.2) - *CC BY 4.0*

◦ 16: Vitamin D (Chapter 18b) - *CC BY 4.0*
- 16.1: Functions of vitamin D (18b.1) - *CC BY 4.0*
- 16.2: Metabolism of vitamin D (18b.2) - *CC BY 4.0*
- 16.3: Vitamin D deficiency in humans (18b.3) - *CC BY 4.0*
- 16.4: Food sources and dietary intakes (18b.4) - *CC BY 4.0*
- 16.5: Nutrient reference values for vitamin D (18b.5) - *CC BY 4.0*
- 16.6: Biomarkers of vitamin D status (18b.6) - *CC BY 4.0*
- 16.7: Interpretive criteria (18b.8) - *CC BY 4.0*
- 16.8: Measurement of serum 25-hydroxyvitamin D (18b.9) - *CC BY 4.0*
- 16.9: Measurement of serum 1,25-dihyroxyvitamin D (18b.10) - *CC BY 4.0*
- 16.10: Other vitamin D metabolites (18b.11) - *CC BY 4.0*
- 16.11: Serum parathyroid hormone (18b.13) - *CC BY 4.0*
- 16.12: Calcium and phosphorus in serum and urine (18b.14) - *CC BY 4.0*

◦ 17: Vitamin C (Chapater 19) - *CC BY 4.0*
- 17.1: Vitamin C (19.1) - *CC BY 4.0*
- 17.2: Functions of Vitamin C (19.2) - *CC BY 4.0*
- 17.3: Absorption and metabolism of vitamin C (19.3) - *CC BY 4.0*
- 17.4: Deficiency of vitamin C in humans (19.4) - *CC BY 4.0*
- 17.5: Food sources and dietary intakes of vitamin C (19.5) - *CC BY 4.0*
- 17.6: Nutrient reference values for vitamin C (19.6) - *CC BY 4.0*
- 17.7: Indices of vitamin C status (19.7) - *CC BY 4.0*
- 17.8: Plasma ascorbic acid (19.8) - *CC BY 4.0*
- 17.9: Ascorbic acid in leukocytes and specific cell subsets (19.9) - *CC BY 4.0*
- 17.10: Ascorbic acid in erythrocytes and whole blood (19.10) - *CC BY 4.0*
- 17.11: Urinary excretion of ascorbic acid and metabolites (19.11) - *CC BY 4.0*
- 17.12: Ascorbic acid in other cells and fluids (19.12) - *CC BY 4.0*
- 17.13: Body pool size (19.13) - *CC BY 4.0*
- 17.14: Functional tests of vitamin C status (19.14) - *CC BY 4.0*

◦ 18: Thiamine (Chapter 20a) - *CC BY 4.0*
- 18.1: Thiamine (20a.1) - *CC BY 4.0*
- 18.2: Biochemical indices of thiamine status (20a.2) - *CC BY 4.0*
- 18.3: Conclusions (20a.3) - *CC BY 4.0*

◦ 19: Riboflavin (Chapter 20b) - *CC BY 4.0*
- 19.1: Riboflavin (20b.1) - *CC BY 4.0*
- 19.2: Biomarkers of riboflavin status (20b.2) - *CC BY 4.0*

◦ 20: Niacin (Chapter 20c) - *CC BY 4.0*
- 20.1: Niacin - Introduction (20c.1) - *CC BY 4.0*
- 20.2: Indices of niacin status (20c.2) - *CC BY 4.0*

◦ 21: Folate (Chapter 22a) - *CC BY 4.0*
- 21.1: Folate (22a.1) - *CC BY 4.0*
- 21.2: Biomarkers of folate status (Chapter 22a.2) - *CC BY 4.0*
- 21.3: Conclusions (22a.3) - *CC BY 4.0*

◦ 22: Vitamin B12 (Chapter 22b) - *CC BY 4.0*
- 22.1: Introduction (22b.1) - *CC BY 4.0*
- 22.2: Biomarkers of vitamin B12 status (22b.2) - *CC BY 4.0*

◦ 23: Calcium (Chapter 23a) - *CC BY 4.0*
- 23.1: Calcium (23a.1) - *CC BY 4.0*
- 23.2: Functions of calcium (23a.2) - *CC BY 4.0*
- 23.3: Absorption and metabolism of calcium (23a.3) - *CC BY 4.0*
- 23.4: Dietary sources and intakes of calcium (23a.4) - *CC BY 4.0*
- 23.5: Nutrient reference values for calcium (23a.5) - *CC BY 4.0*
- 23.6: Calcium deficiency (23a.6) - *CC BY 4.0*
- 23.7: Effects of high intakes of calcium (23a.7) - *CC BY 4.0*
- 23.8: Biochemical indices of calcium status (23a.8) - *CC BY 4.0*
- 23.9: Measurement of bone mineral content and bone mineral density (23a.9) - *CC BY 4.0*

◦ 24: Phosphorus (Chapter 23b) - *CC BY 4.0*
- 24.1: Phosphorus (23b.1) - *CC BY 4.0*
- 24.2: Biological forms of phosphate and their measurement (23b.2) - *CC BY 4.0*
- 24.3: Interpretive criterion- serum phosphate (23b.3) - *CC BY 4.0*
- 24.4: Phosphate balance (23b.4) - *CC BY 4.0*
- 24.5: Hormonal regulation of serum phosphate (23b.5) - *CC BY 4.0*
- 24.6: Nutrient reference values (23b.6) - *CC BY 4.0*
- 24.7: Dietary sources - *CC BY 4.0*
- 24.8: Abnormalities in serum phosphate (23b.8) - *CC BY 4.0*
- 24.9: Inherited and tumor-induced disorders of phosphate metabolism - *CC BY 4.0*

◦ 25: Phosphorus (Chapter 23b) - *CC BY 4.0*
- 25.1: Introduction (23c.1) - *CC BY 4.0*

- 25.2: Functions of magnesium (23c.2) - *CC BY 4.0*
- 25.3: Absorption and metabolism (23c.3) - *CC BY 4.0*
- 25.4: Magnesium deficiency in humans - *CC BY 4.0*
- 25.5: Food sources and dietary Intakes - *CC BY 4.0*
- 25.6: Effects of high magnesium intakes (23c.6) - *CC BY 4.0*
- 25.7: Occurrence of dietary deficiency (23c.7) - *CC BY 4.0*
- 25.8: Indices of magnesium status (23c.8) - *CC BY 4.0*

◦ 26: Magnesium (Chapter 23c) - *CC BY 4.0*
- 26.1: Introduction (23c.1) - *CC BY 4.0*
- 26.2: Functions of magnesium (23c.2) - *CC BY 4.0*
- 26.3: Absorption and metabolism (23c.3) - *CC BY 4.0*
- 26.4: Magnesium deficiency in humans (23c.4) - *CC BY 4.0*
- 26.5: Food sources and dietary Intakes (23c.5) - *CC BY 4.0*
- 26.6: Effects of high magnesium intakes (23c.6) - *CC BY 4.0*
- 26.7: Occurrence of dietary deficiency (23c.7) - *CC BY 4.0*
- 26.8: Indices of magnesium status (23c.8) - *CC BY 4.0*

◦ 27: Copper (Chapter 24b) - *CC BY 4.0*
- 27.1: Copper (24b.0) - *CC BY 4.0*
- 27.2: Serum copper (24b.1) - *CC BY 4.0*
- 27.3: Serum ceruloplasmin (24b.2) - *CC BY 4.0*
- 27.4: Erythrocyte superoxide dismutase (24b.3) - *CC BY 4.0*
- 27.5: Cytochrome c oxidase (24b.4) - *CC BY 4.0*
- 27.6: Skin lysyl oxidase (24b.5) - *CC BY 4.0*
- 27.7: Diamine oxidase (24b.6) - *CC BY 4.0*
- 27.8: Copper chaperones (24b.7) - *CC BY 4.0*
- 27.9: Other indices of copper status (24b.8) - *CC BY 4.0*

◦ 28: Zinc (Chapter 24c) - *CC BY 4.0*
- 28.1: Zinc (24c.1) - *CC BY 4.0*
- 28.2: Dietary Zinc Intake (24c.2) - *CC BY 4.0*
- 28.3: Serum zinc (24c.3) - *CC BY 4.0*
- 28.4: Linear growth (24c.4) - *CC BY 4.0*
- 28.5: Urine zinc (24c.5) - *CC BY 4.0*
- 28.6: Hair zinc concentrations (24c.6) - *CC BY 4.0*
- 28.7: Nail zinc (24c.7) - *CC BY 4.0*
- 28.8: Metallothionein concentrations in circulating blood cell types (24c.8) - *CC BY 4.0*
- 28.9: Metallothionein mRNA expression in circulating blood cell types (24c.9) - *CC BY 4.0*
- 28.10: Zinc transporter mRNA expression in circulating blood cell types (24c.10) - *CC BY 4.0*
- 28.11: Oxidative stress and DNA integrity (24c.11) - *CC BY 4.0*
- 28.12: Kinetic markers: pool sizes and turnover rates (24c.12) - *CC BY 4.0*
- 28.13: Taste acuity (24c.13) - *CC BY 4.0*
- 28.14: Zinc-dependent enzymes (24c.14) - *CC BY 4.0*
- 28.15: Erythrocyte zinc (24c.15) - *CC BY 4.0*
- 28.16: Leukocyte zinc (24c.16) - *CC BY 4.0*
- 28.17: Salivary zinc (24c.17) - *CC BY 4.0*
- 28.18: Multiple biomarkers (24c.18) - *CC BY 4.0*

◦ 29: Selenium (Chapter 25b) - *CC BY 4.0*
- 29.1: Overview (25b.1) - *CC BY 4.0*
- 29.2: Selenium Chemistry (25b.2) - *CC BY 4.0*
- 29.3: Health Effects (25b.3) - *CC BY 4.0*
- 29.4: Selenium Requirements (25b.4) - *CC BY 4.0*
- 29.5: Food Sources (25b.5) - *CC BY 4.0*
- 29.6: Enteric Absorption (25b.6) - *CC BY 4.0*
- 29.7: Metabolism (25b.7) - *CC BY 4.0*
- 29.8: Biochemical Functions (25b.8) - *CC BY 4.0*
- 29.9: Assessing Selenium Status (25b.9) - *CC BY 4.0*
- 29.10: Questions Remain (25b.10) - *CC BY 4.0*

◦ Back Matter - *CC BY 4.0*
- Index - *CC BY 4.0*
- Glossary - *CC BY 4.0*
- Detailed Licensing - *Undeclared*

This page titled Principles of Nutritional Assessment 3e (Gibson et al.) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rosalind S. Gibson et al..