Contents

Development of an international growth standard for preadolescent and adolescent children

Nancy F. Butte and Cutberto Garza, guest editors

Food and Nutrition Bulletin, vol. 27, no. 4 (supplement)

United Nations University Press
Published by the International Nutrition Foundation for The United Nations University
53-70 Jingumae 5-chome, Shibuya-ku, Tokyo 150-8925, Japan
Tel.: (03) 3499-2811 Fax: (03) 3406-7345
E-mail: mbox@hq.unu.edu
ISSN 0379-5721
Design and production by Digital Design Group, Newton, MA USA
Printed on acid-free paper by Webcom, Toronto, ON Canada

Evaluation of the feasibility of international growth standards for school-aged children and adolescents

Nancy F. Butte, Cutberto Garza, and Mercedes de Onis

Abstract

Development of an international growth standard for the screening, surveillance, and monitoring of school-aged children and adolescents has been motivated by two contemporaneous events: the global surge in childhood obesity and the release of a new international growth standard for infants and preschool children by the World Health Organization (WHO). If a prescriptive approach analogous to that taken by WHO for younger children is to be adopted for school-aged children and adolescents, several issues would have to be addressed regarding the universality of growth potential across populations and how to define optimal growth in children and adolescents. A working group concluded that subpopulations exhibit similar patterns of growth when exposed to similar external conditioners of growth. However, on the basis of available data, it cannot be ruled out that some of the observed differences in linear growth across ethnic groups reflect true differences in genetic potential rather than environmental influences. Therefore, the sampling frame for the development of an international growth standard for children and adolescents would have to include multiethnic sampling strategies designed to capture the variation in human growth patterns. A single international growth standard for school-aged children and adolescents could be developed with careful consideration of the population and individual selection criteria, study design, sample size, measurements, and statistical modeling of primary growth and secondary ancillary data. The working group agreed that existing growth references for school-aged children and adolescents have shortcomings, particularly for assessing obesity, and that appropriate growth standards for these age groups should be developed for clinical and public health applications.

Key words: Body-mass index, growth references, height, weight

Introduction

Development of an international growth standard for the screening, surveillance, and monitoring of school-aged children and adolescents has been motivated by two contemporaneous events: the global surge in childhood obesity [1] and the release of a new international growth standard for infants and preschool children [2]. Recognition of the limitations of existing growth references used for assessing childhood obesity (e.g., the National Center for Health Statistics/World Health Organization (NCHS/WHO) growth reference [3], the Centers for Disease Control and Prevention (CDC) 2000 growth charts [4], and the International Obesity Task Force (IOTF) cutoff points [5]) and the release of a new growth standard for infants and preschool children by WHO, in collaboration with the United Nations University (UNU) and other UN agencies, governments, and nongovernmental organizations, have created the urgency and desirability of harmonizing growth assessment tools conceptually and pragmatically. This new growth standard for infants and toddlers was developed from the Multicentre Growth Reference Study (MGRS) and released in April 2006. The MGRS was designed to describe how children

Nancy F. Butte is affiliated with the USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA; Cutberto Garza is affiliated with Boston College, Chestnut Hill, Massachusetts, USA; Mercedes de Onis is affiliated with the Department of Nutrition for Health and Development, World Health Organization, Geneva.

Please direct queries to the corresponding author: N. F. Butte, Children's Nutrition Research Center, 1100 Bates Street, Houston, TX 77030; e-mail: nbutte@bcm.edu.

should grow rather than how they grew in a particular time and place [6, 7]. In part, the prescriptive approach on which the new standard was based required an expanded definition of "health," one that went beyond the absence of overt disease to the adoption of lifestyle practices that support optimal growth and development. If a prescriptive approach analogous to that taken by WHO and UNU for younger children is to be adopted for the development of a growth standard for school-aged children and adolescents, several issues would have to be addressed regarding the universality of growth potential across populations and the characteristics of children and adolescents who are most likely to exhibit optimal growth.

Methods

This paper reviews the feasibility of developing a single international growth standard, with height, weight, and body-mass index (BMI) as primary measures for school-aged children and adolescents based on the proceedings of a meeting convened in Geneva on 16–19 January 2006 by the UNU Food and Nutrition Program in collaboration with the WHO Department of Nutrition for Health and Development and the Nutrition Department of the Food and Agriculture Organization (FAO). At this meeting, 11 position papers were presented and discussed. They included theoretical, biological, and pragmatic issues pertaining to the development of an international growth reference for school-aged children and adolescents with representatives from WHO, CDC, UNU, FAO, the World Food Program (WFP), UNICEF, IOTF, and the International Association for Study of Obesity (IASO).

In this paper, issues pertaining to the feasibility of developing an international growth standard for school-aged children and adolescents will be evaluated, drawing from the 11 position papers [3, 8–17]. The first issue to be addressed is whether it is possible to develop a single international growth reference for children over 5 years of age that is representative and useful for the global population, given possible genetic differences in growth potential across populations. The second issue is whether a prescriptive approach can be taken to develop a growth standard for school-aged children and adolescents from historical growth data, prospective growth data, or both.

Results and discussion

Important limitations of the NCHS/WHO growth reference warrant its replacement with a truly international reference. The current reference was based on the 1977 NCHS growth charts [18]. Age-sex specific BMI percentiles [19] based on 1971–74 National Health and Nutrition Examination Survey I (NHANES I) data were endorsed for global use by WHO [20]. The NCHS/WHO reference was developed from cross-sectional data collected from four separate samples of children and adolescents surveyed in the United States between 1963 and 1974. Unfortunately, the NCHS/WHO reference may not describe optimal growth, given the extent of its positive skewness in body weight, a drawback shared by other, more recent references such as the CDC 2000 reference and the IOTF cutoff points [21]. The upward skewness of these three references results in a substantial underestimation of obesity in school-aged children and adolescents [21–25].

In 1995, a WHO Expert Committee outlined several desirable features for the development of a new international growth reference [20]. The sample should represent healthy children undergoing unconstrained, but not excessive, growth from several developing and developed countries. Secular trends in growth should be small or absent in the sampled population. The sample size should be sufficient to reflect normal variance and to permit an estimation of the more extreme percentiles of weight and height distributions. Cutoff points for under- and overweight should be derived in terms of specificity, sensitivity, and positive predictive values of functional and health-related outcomes.

The technical shortcomings of existing growth references are addressable, but the fundamental issue challenging the development of a single international growth reference is the legitimacy of combining subpopulations, given possible genetic differences in growth potential. The universality of human growth was demonstrated for preschool-aged children reared under favorable nutritional and environmental conditions, regardless of genetic or ethnic background [26], as described by Martorell and Habicht [27]. The feasibility of developing a single international growth reference was challenged by Eveleth and Tanner [28], who pointed to differences in achieved height and growth patterns across subpopulations of children and adolescents.

To expand on studies tabulated in *Worldwide Variation in Human Growth,* by Eveleth and Tanner [28], Haas and Campirano reviewed the literature since 1988 on interpopulation variation in achieved height [9]. Growth data of nominally healthy, privileged children across five major geographic regions—Africa, East Asia, South Asia, West Asia, and Europe—were compiled and compared against the NCHS/WHO reference. Multiracial immigrants moving to more advantaged environments were also included. The major findings from this latest review were as follows. African children and adolescents of upper socioeconomic status achieved similar heights to the NCHS/WHO reference medians, although studies are few. African-American boys and girls achieved or exceeded the median values. The mean heights achieved by East Asian children

and adolescents were below the NCHS/WHO median at all ages from 7 to 18 years, except for recent values from Beijing [29] and Taiwan [30]. In these studies, heights were similar to the NCHS/WHO reference until puberty, at which time mean heights fell to about the 25th percentile. Similarly, heights of boys and girls from South and West Asia tended to follow or were slightly below the NCHS/WHO reference until the age of 11 to 13 years, at which time they fell to about 5 cm below the reference. The heights of children from central Europe tended to be 2 to 4 cm less than the NCHS/WHO median, whereas the heights of children from southern and northern Europe tended to be similar to the NCHS/WHO median. At puberty, the mean heights of children from most European populations approached the reference median, except for adolescents from northern Europe, where heights were 4 to 7 cm above the reference at 18 years.

Even though the above studies focused on nominally healthy, privileged children, secular trends in linear growth still may be occurring in some of the regions. Therefore, Haas et al. [9] compared the tallest children from various ethnic or geographic regions, who presumably attained their genetic potential in linear growth. The mean heights of these boys and girls tracked along the median of the NCHS/WHO reference, with an average difference of about 5 cm between the ages 7 and 13 years. By 15 years, the mean heights of Mexican-American and Japanese adolescents fell to about 5 cm below and the Dutch means increased to approximately 5 to 7 cm above the reference. The mean heights of children living under privileged conditions worldwide did not vary by more than 4 cm from 7 years of age until the initiation of puberty. During adolescence, the mean heights in all populations, except those of European origin, were about 5 to 6 cm (about 0.6 SD) below the NCHS/WHO reference median, and those from northern Europe exceeded the reference median by 1.0 SD at 18 years of age. It remains to be determined whether these differences in adolescent linear growth for non-European populations represent full attainment or some unrealized gain in genetic potential. Whether the degree of geographic isolation and ancestral environmental exposures experienced by some subpopulations are sufficient to affect the genes that control linear growth is unknown. If subpopulation differences in height achieved under optimal environmental conditions persist, genetic differences in growth potential may be responsible.

In general, growth parameters, including height and weight, are highly heritable traits [10]. Determinants of human growth, such as the timing and tempo of puberty and other measures of skeletal and sexual maturation, are also largely under genetic control. Weight, fat mass, and fat distribution are influenced to a larger extent by environmental factors, although genetic factors also are significant. Heritability estimates for growth parameters are lower in nonaffluent populations, probably because of the more pronounced influence of specific nongenetic factors, such as disease and nutrition, in those populations. Limited data are available on cross-population effects of specific genes or gene variants on growth during childhood and adolescence. Genetic epidemiologic studies are needed in different regions of the world to better explore population differences in gene frequencies and gene–environment interactions. Although the fundamental genetic underpinnings of human growth are likely to be essentially the same worldwide, the frequencies of allelic gene variants and gene–environment interactions that influence growth and maturation may differ across populations. Their relative influence in different groups, however, remains unknown.

In the development of a single international growth standard, average growth and normal variation in growth across populations must be represented. This should not pose an insurmountable problem, since the largest variance in complex traits such as weight and height is usually contained within any sufficiently large sample of children from any given population.

There was consensus in the working group that humans follow a similar pattern of growth across ethnic groups and geographic locations. When exposed to similar external conditioners of growth, subpopulations exhibit similar patterns of growth. This was demonstrated years ago for children under 5 years of age [26] and was more recently confirmed by the WHO Multicentre Growth Reference Study [31]. Although the data for children 5 years of age or older are more limited, similar growth patterns across subpopulations were accepted as a general principle by this working group. Therefore, it was concluded that a single standard can describe universal human growth patterns. However, based on available data, it cannot be ruled out that some of the observed differences in linear growth across ethnic groups reflect true differences in genetic potential rather than the sole influence of environmental factors. Therefore, the sampling frame for the development of an international growth standard for children and adolescents would have to include multiethnic sampling strategies designed to capture the variation in human growth patterns.

In addressing the issue of the feasibility of adopting a prescriptive approach to develop a new international growth standard for school-aged children and adolescents from historical growth data, prospective growth data, or both, it is useful to reaffirm the operational difference between *growth references* and *growth standards*. A reference describes a growth pattern of a defined population, whereas a standard defines a recommended pattern of growth that has been associated empirically with specified health outcomes and the minimization of long-term risks of disease.

In the short term, a cross-sectional growth reference

that approaches a standard for universal use could be constructed by using carefully selected historical datasets that reflect realized growth potential and good health of school-aged children and adolescents [11]. The reference population should be one that has stabilized in terms of secular increments in height and weight and that has not been subjected to discernible external constraints on growth (dietary deficiencies, infections, etc.). But since historical datasets seldom have detailed subject descriptors, the health status of the cohort would be unqualified. Furthermore, the available datasets are not representative of the global population, and therefore, such an interim cross-sectional growth reference should be viewed as provisional.

In the long term, a mixed longitudinal growth standard reflective of the multiethnic populations across the regions of the world could be developed. By using prospective data to develop an international growth standard, a prescriptive approach is possible if careful consideration is given to selecting populations or subgroups living in communities that support healthy lifestyles and thereby, presumably, optimal growth. Thus, communities sampled for the development of a growth "standard" would not be representative of national or regional populations but would be uniquely defined on the basis of broadened criteria for health in a manner analogous to that of the new WHO growth standard for infants and preschool children [6]. Criteria that specify healthy behaviors at the individual level would be applied to generate prescriptive-based data. It is very important that the samples selected be free from obesity as well as constrained growth.

A mixed longitudinal design would produce the most useful growth data in the shortest period of time [12]. The sampling frame for the development of a prescriptive growth standard would involve identification of a given number of countries that are broadly representative of the global community, drawing samples of children who are subject to inclusion and exclusion criteria to ensure unconstrained but not excessive growth. The sample size depends on the complexity of the growth curve; prior to puberty, the growth patterns for height and weight are relatively simple, in contrast to the more complex pattern of the pubertal growth spurt, which would require a larger sample size.

Major environmental influences on the growth of children and adolescents must be considered for the selection of individuals and populations in the development of an international growth standard [13, 32]. Inclusion criteria should encompass adequate nutrition, lack of significant endemic rates of infection, and socioeconomic status that does not constrain growth. Low birthweight, catch-up growth, breastfeeding, and early adiposity rebound can affect growth and/or body composition into puberty. Exclusion criteria might include low birthweight due to identifiable pathologies and catch-up growth for individuals, and high altitude and exposure to extremely high levels of environmental pollution for populations. Populations with minimal evidence of secular trends in growth should be chosen. Positive secular trends have been documented in European, European-origin, and Asian populations where mean heights and weights across generations have been shown to be greater, while the sexual maturation and adolescent growth spurt have taken place at progressively younger ages. The average secular increase in height in Europe and North America between 1880 and 1980 was more pronounced during adolescence because of the tempo effect (2–3 cm per decade) and less so during childhood (1–2 cm per decade) [28]. In Japan between 1950 and 1980, the secular trend in height was almost entirely due to the increase in length of the legs. The age of menarche decreased during the last century by about 3 to 4 months per decade in most European countries. In Japan between 1950 and 1975, there was a dramatic decline in the age of menarche in the general population of approximately 1.0 years per decade; the age of menarche in Japan now is as early as or earlier than that in the majority of European populations. Negative secular trends also have been seen among populations in Africa, Papua New Guinea, and Latin America that are largely attributable to socioeconomic and political deterioration; populations under such psychosocial stress should be excluded from the sampling frame.

Since biological maturation is closely related to growth, indicators of biological maturation, including sexual, skeletal, morphological, and/or dental maturity, must be included in the data collection for the development of a growth standard [14]. The timing, sequence, and tempo among maturity indicators of growth must be considered. Maturation of the skeleton is usually monitored with standardized radiographs, and assessment of maturity is based on changes occurring from initial ossification to the attainment of adult morphology of the bones of the hand and wrist. Sexual maturation begins with early embryonic differentiation and ends with full maturity of the sexual organs and fertility. Assessment of sexual maturation is based on secondary sex characteristics—breast development, pubic hair, and menarche in girls, and genital development and pubic hair in boys. Ratings can be performed by clinical examination or self-examination with the use of standardized drawings. In European and North American girls, the average age at takeoff of the adolescent growth spurt is between 8.0 and 10.3 years, and peak height velocity (PHV) occurs two years later (10.8 to 12.2 years) Maturation events occur two years later in boys. Interindividual variation within populations is considerable. Indicators of sexual maturation, age at takeoff of the growth spurt, age at PHV, and skeletal maturity are recommended indicators of the maturation process.

Direct methods for the determination of size and

structure, including height, weight, skinfold thicknesses, and waist circumference, are well established and can be used to monitor linear growth, body mass, ponderosity, abdominal fat, and fat distribution [8]. More complex body-composition methods, such as dual x-ray absorptiometry (DXA) and hydrometry, would be desirable, given the mounting evidence of the relationship of body fat to the risk of cardiovascular diseases and diabetes in children and adolescents.

Measurement of physical activity and physical fitness as indicators of a healthy lifestyle should be incorporated in the development of international growth standards for children and adolescents [15]. Physical activity plays an important role in the regulation of weight, fat mass, and the structural and functional integrity of bone and skeletal muscle, but probably not height or the maturation process. Physical fitness changes with age, growth, and maturation independently of physical activity. Physical activity is assessed by questionnaires, interviews, diaries, direct observation, film or video, devices such as pedometers and accelerometers for measuring movement, heart rate monitoring, measurement of oxygen consumption, and doubly labeled water for the assessment of energy expenditure. Commonly used indicators of physical fitness are cardiorespiratory endurance (endurance shuttle run), function of the lower back (strength and flexibility), and many health-related fitness batteries. Although data for children and adolescents are limited, they suggest a relationship, though moderate, between physical activity and fitness and a favorable risk profile.

In conclusion, an international growth standard for school-aged children and adolescents could be constructed with careful consideration of the population and individual selection criteria, study design, sample size, measurements, and statistical modeling of primary growth and secondary ancillary data. The working group agreed that the NCHS/WHO growth reference for school-aged children and adolescents, the CDC 2000 growth charts, and the IOTF cutoff points all have shortcomings and that a more appropriate growth standard for these age groups should be developed for use in clinical and public health applications.

References

1. Lobstein T, Baur L, Uauy R; IASO International Obesity TaskForce. Obesity in children and young people: a crisis in public health. Obes Rev 2004;5(suppl 1):4–104.
2. World Health Organization Multicentre Growth Reference Study Group. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. Geneva: WHO, 2006.
3. Wang Y, Moreno LA, Caballero B, Cole TJ. Limitations of the current World Health Organization growth references for children and adolescents. Food Nutr Bull 2006;27(suppl):S175–88.
4. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, Wei R, Curtin LR, Roche AF, Johnson CL. 2000 CDC growth charts for the United States: methods and development. Vital Health Stat 11 2002;(246):1–190.
5. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240–3.
6. de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: planning, study design, and methodology. Food Nutr Bull 2004;25(suppl 1):S15–26.
7. Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. Rationale for developing a new international growth reference. Food Nutr Bull 2004;25(suppl 1):S5–14.
8. Lohman TG, Going SB. Body-composition assessment for development of international growth standards for preadolescent and adolescent children. Food Nutr Bull 2006;27(suppl):S314–25.
9. Haas JD, Campirano F. Interpopulation variation in height among children 7 to 18 years of age. Food Nutr Bull 2006;27(suppl):S212–23.
10. Thomis MA, Towne B. Genetic determinants of prepubertal and pubertal growth and development. Food Nutr Bull 2006;27(suppl):S257–78.
11. Seidell JC, Doak CM, de Munter JSL, Kuijper LDJ, Zonneveld C. Cross-sectional growth references and implications for the development of an international growth standard for school-aged children and adolescents. Food Nutr Bull 2006;27(suppl):S189–98;
12. Cole TJ. The international growth standard for school-aged children and adolescents: Statistical considerations. Food Nutr Bull 2006;27(suppl):S237–43.
13. Ulijaszek SJ. The International Growth Standard for Children and Adolescents project: Environmental influences on preadolescent and adolescent growth in weight and height. Food Nutr Bull 2006;27(suppl):S279–94.
14. Beunen GP, Rogol AD, Malina RM. Indicators of biological maturation and secular changes in biological maturation. Food Nutr Bull 2006;27(suppl):S244–56.
15 Malina RM. Physical activity and fitness in an international growth standard. Food Nutr Bull 2006;27(suppl): S295–313.
16. Pelletier D. Theoretical considerations related to cutoff-points. Food Nutr Bull 2006;27(suppl):S224–36.
17. Himes JH. Long-term longitudinal studies and implications for the development of an international growth reference for children and adolescents. Food Nutr Bull 2006; 27(suppl):S199–211.
18. Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF. NCHS growth curves for children birth–18 years. United States. Vital Health Stat 11 1977;(165):i–iv, 1–74.
19. Must A, Dallal GE, Dietz WH. Reference data for

obesity: 85th and 95th percentiles of body mass index (wt/ht^2) and triceps skinfold thickness. Am J Clin Nutr 1991;53:839–46.
20. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser 1995;854:1–452.
21. de Onis M. The use of anthropometry in the prevention of childhood overweight and obesity. Int J Obes Relat Metab Disord 2004;28(suppl 3):S81–5.
22. de Onis M, Dasgupta P, Saha S, Sengupta D, Blossner M. The National Center for Health Statistics reference and the growth of Indian adolescent boys. Am J Clin Nutr 2001;74:248–53.
23. Kain J, Uauy R, Vio F, Albala C. Trends in overweight and obesity prevalence in Chilean children: comparison of three definitions. Eur J Clin Nutr 2002;56:200–4.
24. Vignerova J, Lhotska L, Blaha P. Proposed standard definition for child overweight and obesity. Cent Eur J Publ Health 2001;9:145–616.
25. Fu WPC, Lee HC, Ng CJ, Tay YK, Kau CY, Seow CJ, Siak JK, Hong CY. Screening for childhood obesity: international vs population-specific definitions. Which is more appropriate? Int J Obes Relat Metab Disord 2003;27:1121–6.
26. Habicht JP, Martorell R, Yarbrough C, Malina RM, Klein RE. Height and weight standards for preschool children. How relevant are ethnic differences in growth potential? Lancet 1974;1(7858):611–4.
27. Martorell R, Habicht J-P. Growth in early childhood in developing countries. In: Falkner F, Tanner JM, eds. Human growth: a comprehensive treatise. 2nd ed. New York: Plenum Press, Vol. 3;1986:241–62.
28. Eveleth P, Tanner JM. Worldwide variation in human growth. 2nd ed. Cambridge, UK: Cambridge University Press, 1991.
29. Li H, Leung SS, Lam PK, Zhang X, Chen XX, Wang SL. Height and weight percentile curves of Beijing children and adolescents 0–18 years, 1995. Ann Hum Biol 1999;26:457–71.
30. Chen JY, Chang H, Pan WH. A modified locally weighted method for developing reference standards for height, weight, and body mass index of boys and girls aged 4 to 18 in Taiwan. Hum Biol 2003;75:749–70.
31. WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:56–65.
32. Bogin B. Cambridge studies in biological and evolutionary anthropology. Patterns of human growth. 2nd ed. Cambridge, UK: Cambridge University Press, 1999.

Limitations of the current World Health Organization growth references for children and adolescents

Youfa Wang, Luis A. Moreno, Benjamin Caballero, and Tim J. Cole

Abstract

Since the 1970s, the World Health Organization (WHO) has recommended the use of the growth references developed by the United States National Center for Health Statistics (NCHS) based on national survey data collected in the 1960s and 1970s. These references are known as the WHO or NCHS/WHO growth references. Over the past three decades, the WHO or NCHS/WHO growth references have played an important role internationally in the assessment of child and adolescent growth and nutritional status. However, the references have a number of weaknesses. The limitations of the infant portion of the references were thoroughly assessed in WHO's effort to develop a new international growth reference for infants and preschool children. The present report discusses the limitations of the NCHS/WHO references for school-aged children and adolescents, including a number of conceptual, methodological, and practical problems. The global obesity epidemic poses another challenge that the NCHS/WHO reference cannot appropriately meet. There is a need for a single international reference to assess the nutritional status and growth of school-aged children and adolescents across different countries.

Key words: Adolescent, child, growth reference, international, limitation, obesity

Youfa Wang and Benjamin Caballero are affiliated with the Center for Human Nutrition, Department of International Health, Johns Hopkins University, Baltimore, Maryland, USA; Luis A. Moreno is affiliated with the E.U. Ciencias de la Salud, Universidad de Zaragoza, Spain; Tim J. Cole is affiliated with the Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College London, London.

Please direct queries to the corresponding author: Youfa Wang, Center for Human Nutrition, Room E2546, Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205, USA; e-mail: ywang@jhsph.edu.

For the past 30 years, the World Health Organization (WHO) has recommended the use of the growth references developed by the United States National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC), based on national survey data collected in the 1960s and 1970s. These are called the WHO growth references, the NCHS/WHO growth references, or the NCHS/WHO growth chart [1–3]. The limitations of the infant portion of the current NCHS/WHO growth references have been thoroughly assessed in WHO's effort to develop a new international growth reference for young children from birth to 5 years of age [1, 2, 4–7] In 2000 the CDC published new growth charts to replace the old ones for assessing the size and growth patterns of infants, children, and adolescents in clinical practice and research [8, 9]. In 2006, the Department of Nutrition for Health and Development at WHO, the United Nations University Food and Nutrition Program (UNU-FNP), and the Food and Agriculture Organization (FAO) convened a team of experts to assess the feasibility and appropriateness of developing a single international growth reference or standard for school-aged preadolescents and adolescents. As part of the current effort, the present report focuses on school-aged children and adolescents. After a brief introduction to the current WHO growth references, the report addresses the following issues: limitations of the reference population database; methodological limitations; inconsistency between preadolescents and adolescents; racial or ethnic differences and secular trends in growth and maturation patterns; uncertainty and inadequacy for assessing linear growth in adolescents; new challenges and needs posted by the growing global obesity epidemic; and adjustment for between-population maturity differences.

What are the current WHO growth references?

The current WHO growth references (**table 1**) were

TABLE 1. Summary of the current NCHS/WHO growth references[a]

Outcomes	Anthropometric measures and cutoff points	Indication of growth or nutrition problems
Infants and children (< 10 yr)		
Stunting	HAZ < –2, or < 3rd percentile	Chronic malnutrition
Wasting/thinness	WHZ < –2, or < 3rd percentile	Acute malnutrition, current malnutrition
Overweight	WHZ >2	Overweight
Adolescents (≥10 yr)		
Stunting	HAZ < –2, or < 3rd percentile	Chronic malnutrition
Thinness	BMI-for-age < 5th percentile	Underweight
At risk for overweight	BMI-for-age ≥ 85th percentile	Overweight
Obese	BMI-for-age ≥ 85th percentile, and triceps and subscapular skinfold thickness-for-age ≥ 90th percentiles	Obesity

NCHS, National Center for Health Statistics; WHO, World Health Organization; HAZ, height-for-age z-score; WHZ, weight-for-height z-score; BMI, body-mass index

a. According to the WHO/NCHS reference [1, 2] and the BMI percentile developed by Must et al. [18].

recommended by a WHO Expert Committee in 1995 [2] and include two portions: the 1977 NCHS growth charts, which have been previously recommended by WHO [10], and a new set of age- and sex-specific body-mass index (BMI) and skinfold percentiles for adolescent anthropometry, developed from US data. Previously WHO had made no specific recommendation for this age group.

The NCHS growth charts include anthropometric measurements such as weight-for-height, weight-for-age, height-for-age, and head circumference, which were developed on the basis of several datasets (see below). They are generally referred to as the "1977 NCHS growth charts" [11–13]. These include growth charts for infants from birth to 36 months and for older children from 2 to 18 years of age. In 1978, the CDC produced a normalized version of the NCHS curves. WHO subsequently recommended these normalized growth charts for international use in assessing nutritional status in child populations [10, 14–17]. These normalized versions of the 1977 NCHS growth charts are thus sometimes referred to as the NCHS/WHO, CDC/WHO, or NCHS/CDC/WHO growth charts. In 1980, a software version of the reference for mainframe computers was developed by the CDC to facilitate the interpretation of growth data from surveys or clinical studies. Throughout the 1980s, several microcomputer-based software versions of the NCHS/WHO references were developed and supported by the CDC and WHO. These have contributed significantly to the wide acceptance of the concept of the international reference by simplifying the handling of anthropometric data [2]. Details regarding the historical background of the currently used NCHS/WHO growth charts can be found elsewhere [1–3, 14].

In 1995, a WHO Expert Committee reviewed these growth references and new research findings and re-endorsed the use of the 1977 NCHS growth charts [2, 3]. The committee also addressed a number of weaknesses of the NCHS infant growth charts. A major new recommendation by this committee was the use of sex- and age-specific BMI values and triceps and subscapular skinfold thickness percentiles for the classification of obesity in adolescents [2]. These BMI percentiles were developed by Must and colleagues in 1991 on the basis of data from the first National Health and Nutrition Examination Survey (NHANES I) collected in 1971–74 [18], and the skinfold thickness percentiles were developed by others [19, 20]. The committee acknowledged some potential problems in using these US-derived adolescent BMI percentiles, and information regarding their usefulness to predict future risk in children from developing countries is limited. The committee recommended the use of these references on a provisional basis until better reference data for adolescent growth become available [2, 3].

Limitations of the reference population and data

There are conceptual differences between growth "references" and "standards." The WHO Expert Committee defined "reference" as a tool for grouping and analyzing data and for providing a common basis for comparing populations, but not for drawing inferences about the meaning of observed differences. In contrast, a "standard" embraces the notion of a norm or desirable target and thus involves a value judgment [2]. Still, the committee acknowledged that it would be virtually impossible to prevent the use of references as "standards" for judging the nutritional status of individuals and populations. Therefore, the committee pointed out that it is always desirable to choose references that most resemble true standards.

The 1995 WHO Expert Committee defined several desirable features for datasets to be used in the development of a new reference population [2, 3]. These

features include the following: the anthropometric data should represent multiple countries and geographic regions, including less-developed countries; the data should reflect the status of healthy populations with unconstrained growth (even when not representative of the whole population); data for children from birth to adolescence should be included; the sample size and data-collection procedures should be appropriate and well documented, with at least 200 individuals in each age and sex group; data for adolescents should include assessment of sexual developmental stage; and secular trends in growth should be small or absent, because they suggest either constrained or excessive growth or weight gain in the reference population. In light of these recommended characteristics, the NCHS/WHO references have a number of limitations.

First, the current NCHS/WHO references were developed on the basis of data collected in only one country, the United States. As noted by the WHO Expert Committee, including data from several countries will improve the estimate of variability of physiologic growth and will also minimize political concerns that arise from the use of patterns of child growth of a single country as a standard for all other countries. The WHO Working Group on Infant Growth has chosen a prescriptive approach to develop a truly international growth standard for infants and preschool children based on data collected from several countries [2, 4, 6]. It is likely that the US NCHS reference population does not present the optimal growth patterns for all age groups and for all different world regions (see below).

Second, the NCHS reference population was selected by using a *descriptive* approach, which when applied to a population like that of the United States, which has an increasing prevalence of obesity, is likely to result in a nonhealthy sample. For example, the distribution of weights in the NCHS reference is positively skewed, with a long tail to the right and a high prevalence of overweight. According to the NHANES I data for 1971–74, the prevalence of overweight (BMI ≥ 25) was 47.7% in US adults and 15.4% in children aged 6 to 18 years (≥ 85th BMI percentile [21]). The NCHS BMI percentiles for US children are much higher than those of French children. For example, the 85th percentiles for US boys exceeded the 90th percentiles and approximated the 97th percentiles for French boys [2, 3]. This would suggest that the prevalence of overweight among "healthy" US children and adolescents might be higher than 15% if overweight was assessed by using a standard truly representative of a healthy population. Conversely, using 85th BMI percentile from the US charts to classify childhood obesity worldwide would result in many overweight children being classified as lean or of normal weight.

Third, the current WHO reference dataset consists of several unrelated samples, whereas a single survey sample would be more desirable. Several datasets collected in the United States were used, including data from the National Health Examination Survey (NHES II, 1963–65) for the ages of 6 to 11 years, NHES III (1966–70) for the ages of 12 to 17 years, and the First National Health and Nutrition Examination Survey (NHANES I, 1971–74) for the ages of 1 to 18 years. For children under 2 years of age, data from the Fels Longitudinal Study were used. The decision to fit separate curves for the Fels and later data has resulted in a disjunction between the NCHS/WHO references for infants and children and has led to spurious age-related differences in growth status when these references have been used to assess infants and children. These spurious differences, in particular, affect the assessment of growth in height, since length was measured in the Fels study whereas height was measured in the other surveys [1].

Additionally, the Fels dataset reflects the growth of formula-fed rather than breastfed infants. Furthermore, formula composition and feeding practices of 30 years ago may differ substantially from current recommendations [5, 6, 22]. Similar problems may affect the growth references for older children because of the possible long-term effects of the early feeding experience. In addition, the data collected in the 1960s through the 1970s from US children and adolescents may not reflect the desirable eating and growth patterns for these age groups or the more recent patterns worldwide.

Fourth, the NCHS/WHO references were developed on the basis of cross-sectional data and are inadequate for longitudinal growth monitoring for several reasons: the cross-sectional reference does not express growth as a velocity percentile-it gives no clue as to whether or not a given rate of percentile crossing is unusual; there is no adjustment for regression to the mean, whereby smaller children tend to grow faster; and the rapid changes in velocity due to puberty and variability in timing of puberty are not captured by cross-sectional data. Moreover, for most sex and age groups in the NCHS/WHO references, the sample size was approximately 120. For example, the sample sizes used in developing the BMI percentiles ranged between 91 and 153 [18], sizes smaller than the recommended size of 200 [2, 3].

Methodological limitations

Growth assessment parameters—anthropometric indexes

Another limitation of the current NCHS/WHO references is their recommendation of the use of different anthropometric measures and different cutoff points for children and adolescents, particularly in the assessment of overweight and underweight. Age- and sex-specific weight-for-height was used in children (z-score

of 2 for overweight and –2 for underweight) and BMI percentiles in adolescents (85th percentile for "at risk of overweight" and 5th percentile for underweight). These different measures cause a number of conceptual, methodological, and practical problems.

Weight-for-height reference is advantageous in that it does not require knowledge of the individual's chronological age. However, the weight-to-height relationship changes dramatically with age and with maturation status. As a result, at a given height, the weight corresponding to a particular percentile is not the same for all ages, so that the meaning of a given weight-for-height percentile differs according to age [2].

It is recommended that an ideal measure of body fatness should meet several requirements [1, 2], including the following: the measure should be accurate in assessing the amount of body fat; the measure should be precise, with small measurement error; the measure should predict the risks of health consequences; it should be possible to develop some cutoff points to separate individuals into different groups on the basis of their excess adiposity-related health risks; and the measure should be accessible in terms of simplicity, cost, and ease of use and acceptable to the subjects in order to be useful in clinical settings or epidemiologic studies.

Although none of the existing measures satisfies all of these criteria, the current consensus is that BMI is probably the best choice among available measures, including weight-for-height; BMI can be easily assessed at low cost and has a strong association with body fat and health risks. BMI has been recommended for use in children, adolescents, and adults to assess body weight status [8, 23–27].

However, BMI also has a number of limitations as an indirect measure of fatness [28, 29]. One potential limitation is the association of BMI with height in young people, which varies by age and sex [29]. Using data collected in NHANES III for American children aged 2 to 18 years, we calculated the correlation coefficients and the power of height (p) values needed to construct a weight/height index that was not correlated with height (see **fig. 1**). Among boys aged 5 to 16 and girls aged 5 to 11, BMI was correlated with height, and the p values need to be greater than 2 [29]. This confirms the pattern first highlighted by Cole [30, 31]. Franklin [32] presented similar findings of p values for boys aged 6 to 18 from the United States, United Kingdom, Japan, and Singapore. He argued that BMI underestimates the effect of height on weight; taller children tend to have larger BMIs than shorter children, and as a result, they are more likely to be classified as being overweight or obese when age-specific BMI cutoff points are used. Our analyses show that among boys aged 6 to 12 years, tall boys (those in the upper age-specific height tertile) were approximately twice as likely to be overweight as those in the bottom tertile [29, 32].

Some researchers have argued that there is no intrinsic reason why BMI should be uncorrelated with height. Fat children tend to be taller than average, and the higher BMI matches the extra height [33]. Our comparison analysis based on skinfold thickness and BMI suggests that at least some of the height-related differences in the prevalence of overweight defined by BMI cutoff points is due to the association between gains in height and adiposity during late childhood and puberty [29], so that within populations fatter children tend to be taller. What remains to be seen is whether the same relationship holds across populations. The large differences in the heights of young people worldwide (see below) underscore the potential problems of using the US BMI percentiles, as recommended by the WHO Expert Committee. This issue will need to

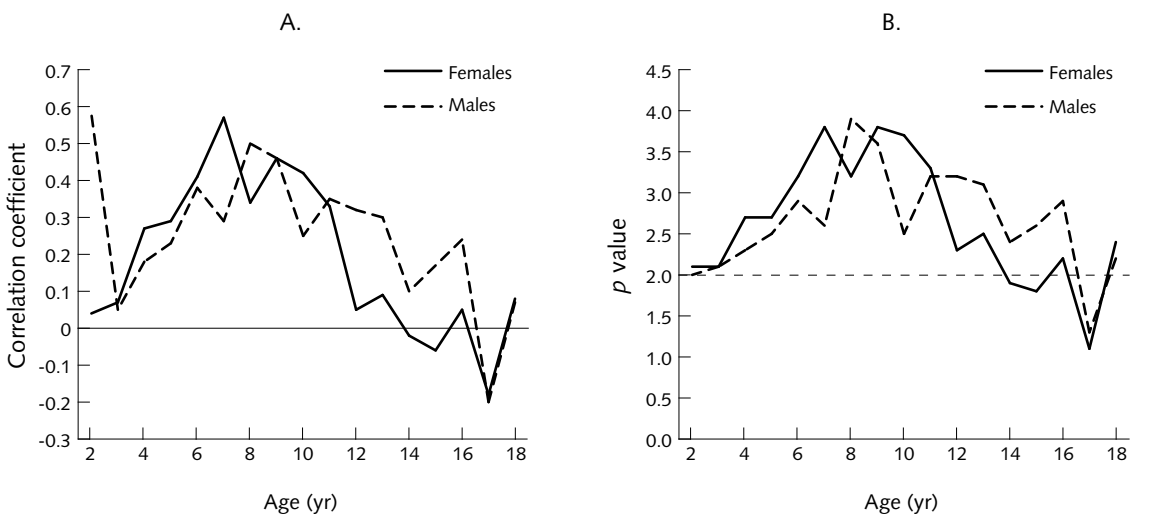

FIG. 1. Correlation coefficients between BMI and height (A) and estimates of the p value in the weight–height p index (B) in American boys and girls, according to age. Data from NHANES III (1988–94). Adapted from Wang [29]

be addressed if, as is likely, BMI will be used in a new international standard.

Studies also show that in comparison with classifications of obesity based on direct measures of body fatness, such as percentage of body fat (%BF), the sensitivities of international BMI cutoff points for screening obese individuals are relatively low, although they have good specificities [34, 35]. For example, Reilly et al. [34] found that when the 95th %BF percentile of the study population distribution was used as the "gold standard" of obesity, the International Obesity Task Force (IOTF) BMI reference (see below) had a very low sensitivity, and it differed markedly between boys and girls (46% versus 72%, respectively). In contrast, the local BMI reference (95th percentile) had a much better sensitivity (88%), although both references had a good specificity (99% versus 94%, respectively). Similarly, on the basis of data collected from children and adolescents in Spain, Sarria et al. [35] found that when the 85th %BF percentile was used as the gold standard, the sensitivity and specificity of BMI cutoff points were approximately 50% to 60% and 85% to 90%, respectively. This suggests that BMI may not be sensitive enough for screening overweight children and adolescents. A more direct measure of body fat, such as skinfold thickness, waist circumference, or bioelectrical impedance, is needed instead.

The WHO Expert Committee recommended that triceps and subscapular skinfold thickness measurements be used in addition to BMI for the assessment of obesity in adolescents, in particular to maximize specificity. In reality, such skinfold thickness measurement data are often not available or are difficult to collect. Although the use of skinfold thickness measurements is logically appealing, since they are "direct" measures of fatness, they pose methodological and practical problems. Skinfold thickness and skinfold compressibility vary according to age, site, sex, and possibly ethnicity [18]. Further, in obese individuals, skinfold thickness measurements have poor reliability [18]. To our knowledge, few researchers have used this WHO obesity reference (i.e., both BMI and skinfold thickness cutoff points).

Emerging evidence suggests that other anthropometric measures might be important and need to be considered in developing new international obesity standards. For example, recent studies predominantly among adults suggest that waist circumference, which is not used in the current NCHS/WHO references, is more closely associated with chronic diseases such as cardiovascular disease and type 2 diabetes [36–38]. Waist circumference may be particularly useful for monitoring changes in individuals in comparison to BMI.

Selection of cutoff points

The NCHS/WHO references, in particular the anthropometric measure cutoff points such as the 2 and –2 z-scores and 85th percentiles, were chosen mainly on the basis of statistical criteria rather than of health outcomes. Ideally, the criteria should be established on the basis of health outcomes, and cutoff points should be chosen at the point most appropriate for the particular purpose in view. The selection of appropriate cutoff points for assessing "high-risk" individuals and population groups should be based on evidence of increased risk of morbidity, mortality, and/or impaired functional performance [1]. However, the reality is that to assess the relationship between different indicators and cutoff points and health outcomes is more difficult in children than in adults. In children, two different types of health outcomes may need to be considered: short- and intermediate-term health outcomes during childhood and adolescence, and long-term health outcomes in adulthood. Well-designed long-term longitudinal studies are needed.

Pelletier [39] suggests that the general approach for selecting cutoff points for assessing growth and nutritional status should be based on three types of considerations: health and functional consequences of deviations in a given anthropometric indicator; differences in age, maturation, sex, ethnicity, and other factors that affect anthropometric measurements independently of, or in conjunction with, the health or social causes or consequences of obesity; and the intended or potential uses of anthropometric indicators, such as clinical diagnosis, policy formulation, social utility, and advocacy for particular problems and solutions. Pelletier argues that on the basis of these considerations, a logical conclusion would be that different indicators and cutoff points are needed for different uses. However, past experience suggests that this conclusion is not likely to be accepted by various user communities, including international expert groups, because of the strong desire to agree upon and promote single, simple indicators and cutoff points for all uses.

Nevertheless, considering our increasing understanding of the complexity of assessing children's and adolescents' growth and the new reality of a growing global obesity epidemic, and considering that many developing countries are facing a double burden of under- and overnutrition problems, the international community and the public might accept complex indicators and cutoff points if appropriate single, simple indicators and cutoff points cannot be developed.

The current NCHS/WHO references use both z-scores and percentiles. In particular, z-scores of –2 and 2 as well as the 97th and 3rd percentiles are used for children, and the –2 z-score as well as the 5th, 85th, and 90th percentiles are used for adolescents. A critical problem is that z-scores of 2 and –2 correspond to

percentiles of 97.7 and 2.3, whereas the 85th and 5th percentiles correspond to z-scores of 1.04 and –1.65, respectively. Thus, these two methods will generate very different estimates of prevalence for children and adolescents. For example, even if two children approximately 10 years of age have similar nutritional statuses based on a gold standard, one may be classified as "normal" and the other as having a nutrition problem simply because of their age difference (e.g., an adolescent aged 10.1 years and a child aged 9.9 years).

The use of z-scores was recommended because of several considerations. First, z-scores are calculated on the basis of the distribution of the reference population (both the mean and the standard deviation); thus, they reflect the reference distribution. Second, as standardized measures, z-scores are comparable across ages, sexes, and indicators [40, 41]. Third, another major advantage is that summary statistics such as means and standard deviations can be calculated from z-scores. In addition, z-score values can quantify the growth status of children outside of the percentile ranges. However, a limitation of z-scores is that they are not easy to explain to the public and may be of limited use in practice. In contrast, percentiles are easily understood and used by both health professionals and the public. The percentile refers to the position of an individual on a given reference distribution. During recent years, a growing consensus among scientists has been to use sex- and age-specific BMI percentiles as cutoff points instead of weight-for-height z-scores for assessing overweight and obesity as well as thinness and underweight in children over 2 years old [1, 2, 8, 23, 24]. However, a limitation of using percentiles is that the same interval of percentile values corresponds to different changes in absolute values in different anthropometric indicators. In addition, the use of percentiles does not allow for quantification of the change in percentile values near the extremes of the reference distribution [2, 42]. For this reason, percentiles should not be used to assess change in status over time; change in z-score is a better measure. Future research needs to examine the associations between anthropometric indicators and cutoff points proposed for assessing growth and long-term health outcomes. The results may support the use of different cutoff points than those developed on the basis of statistical approaches.

Statistical methods and techniques used for curve-fitting and smoothing

Since the development of the NCHS/WHO references, new and better statistical techniques and strategies have been developed for curve-fitting and smoothing. For example, the development of the 2000 CDC growth charts was carried out in two stages: curve-smoothing and transformation. The least mean square (LMS) method that was introduced in the 1980s [43–45] was modified for use in the transformation stage [8, 9]. When the 1977 NCHS/WHO growth curves were developed, a least-squares-cubic-spline technique was used [13]. The NCHS/WHO BMI and skinfold thickness percentiles were developed with the use of LOWESS (LOcally WEighted regression Scatter-plot Smoothing) [18].

Different methods will affect the final curves and cutoff points. For example, the BMI percentiles developed by Must et al. [18] and Hammer et al. [46] on the basis of the NHANES I data are not identical. Take the 5th percentile for white adolescents aged 18 years as an example. According to Must et al., the BMI values for this percentile are 17.5 for males and 16.9 for females; according to Hammer et al., the corresponding values are 18.3 and 17.2. For the 95th percentile, Must et al. found BMI values of 29.9 for males and 29.2 for females; the corresponding values for Hammer et al. are 29.7 and 31.0. Different curve-fitting and smoothing techniques were used in these two studies. In general, methods that summarize the centiles as an underlying distribution, such as the LMS method, "borrow strength" from neighboring ages and centiles and thus make better use of the data than the separate centile-fitting methods used in the NCHS/WHO references.

Inconsistency between preadolescents and adolescents

Different anthropometric parameters, terminology, and statistical cutoff points are used in the current NCHS/WHO references regarding the classifications of overweight, obesity, and underweight for preadolescents (< 10 years) and adolescents (≥ 10 years) (**table 1**). This inevitably leads to inconsistencies of prevalence estimates across the age boundary. A new reference should ensure that the same definitions apply across the age range in order to avoid these problems of inconsistency.

Racial or ethnic differences and secular trends in growth and maturation patterns

Racial or ethnic differences and secular trends in growth, body composition, body build, and sexual maturation are likely to complicate the interpretation of anthropometric measures and cutoff points [28, 47–54]. Some of these problems are addressed in other articles in this issue. An important reason for supporting the use of the US CDC/NCHS growth charts as an international standard (i.e., the NCHS/WHO references) was the consensus that environmental factors rather than genetics are the main determinants of between-population differences in child growth. Many studies of affluent populations have found that the mean

heights of young children differ little across ethnic groups in comparison with the socioeconomic variability within a given ethnic group [55]. Data collected among privileged groups of children in developing countries show that child growth is influenced mainly by socioeconomic status and not by race or ethnicity, and that the distributions of weight-for-height and height-for-age values for the privileged groups and the US CDC/NCHS reference population are nearly identical [15].

It is worth noting that the consensus that environmental factors rather than genetics are the main determinants of between-population differences in child growth applies mainly to young children; however, limited information is available about older groups, and not all researchers share the consensus view. For example, Ashcroft and Desai [56] reviewed data collected from infants and children of African, Indian, Chinese, and European origin in Guyana and Jamaica in order to compare the influence of ethnic origin and environment on anthropometric measurements. They found that the mean heights and weights of African and European children were greater than those of Indian and Chinese children. African children had greater weight-for-height and greater arm circumferences but smaller triceps skinfolds than Indian children. The authors argued that these differences, which could not be explained by nutritional or other environmental causes, indicated that ethnic origins could not be disregarded when assessing nutritional status by anthropometric measurements. It is accepted that there are some variations in the growth patterns among children of different racial or ethnic groups even in developed countries, although these variations are relatively small compared with the large worldwide variation in growth related to health, nutrition, socioeconomic status, and environmental factors [1]. In **table 2** we present the height medians for 10- and 12-year-old boys and girls in several growth references and populations as examples. Some experts in the field have argued against the use of universal international references [65].

A limited number of studies have examined the between-population differences in older children's and adolescents' growth patterns. Growing evidence suggests that there are large between-population differences in their growth and sexual maturation patterns (see other papers by Beunen et al. [66] and by Haas and Campirano [67] in this issue). This is likely to pose even more uncertainty and concern regarding the use of the US NCHS reference in populations other than the US population, in particular populations in developing countries [2].

Evidence is insufficient to disregard the contribution of genetic factors to the between-population differences in growth in adolescents, especially height. Data collected from a group of 818 Bengali boys from middle-class families in India showed that the median height curve of the boys had a similar shape to that of the NCHS/WHO reference but on average was approximately 5 cm lower than the NCHS/WHO reference (ranging from 2.4 cm to 7.8 cm) [68]. The researchers suggested that these differences were the result of a combination of genetic and environmental influences. Another, more recent study conducted among affluent urban Pakistani adolescents aged 10 to 15 years found that the younger adolescents were taller than the US 2000 CDC growth references but the older adolescents were shorter [69].

Some studies have shown differences in young people's height across industrialized countries (see **table 2**), and other data suggest that considerable differences may exist between different ethnic groups even

TABLE 2. Comparison of median heights (centimeters) of 10- and 12-year-old boys and girls in different populations with growth references

	Boys		Girls		
Country or growth reference	10 yr	12 yr	10 yr	12 yr	Source
NCHS/WHO reference	137.5	149.7	138.3	151.5	Hamill et al., 1979 [13]
US 2000 CDC reference	140.1	150.9	139.8	153.4	Kuczmarski et al., 2000 [8]
UK 1990 reference	138.4	148.4	138.4	149.8	Freeman et al., 1995 [57]
China					
National average	132.5	142.5	132.1	143.6	Ge et al., 1995 [58]
High-income group	136.2	146.2	136.3	148.5	
France	135.6	145.9	134.7	147.7	Sempe et al., 1979 [59]
Italy (Central and North)	139.5	151.3	139.6	152.6	Cacciari et al., 2002 [60]
Italy (South)	137.9	149.1	138.6	150.9	
Netherlands	143.2	154.0	143.3	155.3	Fredriks et al., 2000 [61]
Spain	136.5	146.7	136.7	148.4	Hernandez, 1988 [62]
Swedish 1995 reference	140.1	150.5	140.1	153.0	Lindgren et al., 1995 [63]
Swedish 2002 reference	141.1	152.1	141.1	153.9	Wikland et al., 2002 [64]

NCHS, National Center for Health Statistics; WHO, World Health Organization; CDC, Centers for Disease Control and Prevention

within the same country. For example, on the basis of nationwide data collected between 1994 and 2000 from approximately 55,000 Italian schoolchildren aged 6 to 20 years, Cacciari et al. [60] reported that children in central and northern Italy were taller and heavier than their counterparts in the South. At the end of the growth period, the average difference between the Center and North and the South was 2.4 cm for girls and 2.7 cm for boys. The authors argued that these differences in height were unlikely to be due to social, environmental, or nutritional factors. In addition, published data show that age at adiposity rebound varies across populations [29].

Secular changes in growth, body composition, and sexual maturation may complicate the interpretation of anthropometric measures such as weight-for-height and BMI. This poses a challenge for the development of both local and international anthropometric references. For example, Wells et al. [53] showed that for a given BMI value, contemporary Cambridge children have more fat mass and less lean mass than the British reference child. Their findings suggest that BMI-based assessments have underestimated the increase in children's fatness. Changes over time in the relationship between BMI and body composition will give a misleading predicted risk of future adult illness. These changes in the relationship between BMI and health outcomes may make it difficult to justify using fixed age- and sex-specific BMI cutoff points developed from another country such as the United States to assess obesity, particularly in societies that have been experiencing dramatic socioenvironmental changes. Thus, there are two key decisions involved: selection of the database (local data versus data from other places) and choice of cutoff points (e.g., 85th or 90th percentile). However, the reality is that unless each country develops its own reference, which is not feasible or recommended [2], we will inevitably have to use references developed on the basis of data collected from some populations in many other countries.

Uncertainty and inadequacy in assessing linear growth in adolescents

As addressed in previous sections, there are considerable between-population differences in adolescents' linear growth (see **table 2**). It is likely that at least a part of these differences is due to genetic factors, and this has fueled the debates and concerns regarding the use of growth standards developed on the basis of data from one country to assess the linear growth in other populations, in particular in developing countries. Such concerns were expressed in the WHO Expert Committee's 1995 report [2].

A small change in the standard can result in large differences in the prevalence of stunting. For example, Eckhardt and Adair [70] compared the NCHS/WHO and 2000 CDC references for stunting in Filipino children from birth to 16 years. Although the CDC stated that the differences between the two references were minor on the basis of US data, these authors found that the differences in the prevalence of stunting according to the two references were large and inconsistent. For example, the prevalence was approximately 45% on the basis of the NCHS/WHO reference versus 61% on the basis of the 2000 CDC reference in girls aged 8.5 years, 47% versus 37% in girls aged 11.5 years, and 44% versus 48% in girls aged 15 years. On the basis of the NCHS/WHO reference, the prevalence remained relatively consistent between the ages of 8.5 and 15 years, whereas on the basis of the 2000 CDC reference, it decreased. Similarly, Moestue et al. [71] found differences in stunting prevalence among Bangladeshi children depending on whether the NCHS/WHO, CDC 2000, or British 1990 reference was used.

The growing global obesity epidemic

The NCHS/WHO references do not meet the need to address the growing global obesity epidemic. Increasing numbers of studies published worldwide show that the prevalence of overweight and obesity has increased over the past two decades in both developed and developing countries and that we are facing a global epidemic of childhood obesity [72, 73]. Traditionally, the WHO growth references have been developed and widely used with a focus on addressing malnutrition problems. To assess childhood obesity, different classifications and references have been advocated by different organizations and used in different regions and countries (see **table 3**).

For example, recently a number of countries, including the United States, the United Kingdom, France, and China, have developed or re-endorsed their own BMI percentiles for the classification of child and adolescent obesity. These BMI cutoff points differ considerably [8, 75–78], which can result in very different estimates of obesity prevalence. For example, Zimmermann et al. [79] compared the BMI values of a representative national sample of 595 6- to 12-year-old Swiss schoolchildren with the US, UK, French, and Swiss references. Depending on which reference was used, the prevalence of obesity varied between 9.7% and 16.1% and the prevalence of overweight between 21.7% and 34.2%.

The WHO weight-for-height references for defining overweight in children have been used in many previous studies. However, the use of the WHO references for defining overweight and obesity in adolescents has been very limited. Many researchers have chosen instead to use the US 85th and 95th BMI percentiles developed by Must et al. [18] or their local references, such as 120% of ideal body weight. This has made it

TABLE 3. Classifications of child and adolescent overweight and obesity

Classification	Overweight	Obesity	Data and reference population	Source
WHO reference	Child (6–9 yr): WHZ > 2[a] Adolescent (10–18 yr): BMI 85th percentile (called "at risk of overweight")	Child (6–9 yr): no reference Adolescent (10–18 yr): BMI ≥ 85th percentile, and subscapular and triceps skinfolds ≥ 90th percentile	US NHANES I data (1971–74)	WHO, 1995 [2]
IOTF reference	≥ BMI-for-age cutoffs derived from BMI–age curves passed BMI of 25 at age 18	≥ BMI-for-age cutoffs derived from BMI–age curves passed BMI of 30 at age 18	Data from USA, Brazil, Britain, Hong Kong, the Netherlands, and Singapore	Cole et al., 2000 [24]
"Old" US BMI percentiles	≥ BMI 85th percentile (called "at risk of overweight")	≥ BMI 95th percentile (called "overweight")	US NHANES I data (1971–74)	Must et al., 1991 [18]
"New" US BMI percentiles (2000 CDC Growth Chart)	≥ BMI 85th percentile (called "at risk of overweight")	≥ BMI 95th percentile (called "overweight")	US NHANES data (1971–1994)	Kuczmarski et al., 2000 [8]
European–French BMI reference	≥ BMI 90th percentile	≥ BMI 97th percentile	Data from the French population	Poskitt, 1995 [74]; Rolland-Cachera et al., 1991 [75]

WHO, World Health Organization; WHZ, weight-for-height z-score; NHANES, National Health and Nutrition Examination Survey; IOTF, International Obesity Task Force; BMI, body-mass index; CDC, Centers for Disease Control and Prevention.
a. A z-score of 2 corresponds to the 97.7th percentile.

difficult to make international comparisons. In order to develop a universal reference for the classification of childhood obesity, Cole et al. developed a series of sex- and age-specific BMI cutoff points based on data collected in six countries (the United States, Brazil, the United Kingdom, Hong Kong, the Netherlands, and Singapore) and linked to the WHO-recommended BMI cutoff points of 25 and 30 used to define overweight and obesity in adults [24]. These cutoff points differ from the current WHO BMI cutoff points and therefore yield different estimates of the prevalence of obesity [21, 80]. These BMI cutoff points are also called the International Obesity Task Force (IOTF) reference.

Using national data collected from the United States, China, and Russia, we found that if the US BMI 85th percentiles (the NCHS/WHO reference for adolescents) and the IOTF BMI reference were used to define overweight, the prevalence was much lower among adolescents (10–18 years) than children (6–9 years) in China and Russia, but the prevalences among adolescents and children were similar in the United States [21]. We suspect that the age-related difference in prevalence in China and Russia may be due to a possible difference between the sensitivity of these references in different age groups. In other words, these references may be more sensitive in identifying overweight children than overweight adolescents in low- and middle-income countries, for two reasons: growth and development patterns among populations and the BMI–age relationship in developing countries may be different from that in the NCHS/WHO and IOTF reference populations; and between-population differences in sexual maturation status may exist. Children and adolescents from low- and middle-income countries mature later than the reference populations [50, 51, 81].

Therefore, the current WHO reference may not be appropriate for longitudinal assessment of pediatric-age populations. Although a large body of literature shows that, in general, about one-third of obese children and one-half of obese adolescents in the United States and many other industrialized countries remain obese as adults [82, 83], our research and that of Mo-suwan et al. show that on the basis of the WHO and IOTF references, only approximately 10% of overweight children remain overweight as adolescents 5 to 6 years later [84, 85]. We suspect that at least part of the remarked differences is due to a lower sensitivity of the reference in identifying overweight individuals during adolescence than during childhood in developing countries [29, 84].

Use of the NCHS/WHO references may result not only in an underestimation of the obesity problem among older children and adolescents in developing countries, but also in an overestimation of the problem of undernutrition in this age group in these countries. This will cause considerable misclassification of the nutritional status of individual children. In the study

of Indian boys mentioned above, de Onis et al. [68] reported that the NCHS/WHO reference as well as three other European BMI-for-age references yielded an unrealistically high prevalence of underweight. On the basis of the NCHS/WHO references, the prevalence of underweight was 51%, as compared with a low prevalence of stunting of 11.2%.

Recently several researchers have raised concerns about the use of international references for the classification of childhood obesity, similar concerns exist about the WHO references regarding obesity [29, 86, 87]. A dilemma is that according to the current consensus, obesity is considered a disease. Thus, ideally it should be defined primarily on the basis of health consequences and not on the basis of fixed statistical cutoff points. A corollary of this view is that local references have advantages when used for clinical assessment. On the other hand, an appropriate international standard is beneficial for public health use [2]. To develop a new international obesity standard, these issues need to be considered. One possible solution may be to develop an approach for the adjustments of between-population differences and to provide alternatives to serve different goals, as was recently recommended by a WHO Expert Consultation for the classification of obesity in adults in certain populations [88]. Recent evidence suggests that the associations between BMI, percentage of body fat, and health risks are different in Asian and European populations. A substantial proportion of Asian people with BMIs lower than the current NCHS/WHO cutoff point of 25 for overweight are at high risk for type 2 diabetes and cardiovascular disease. The WHO Expert Consultation identified BMI values of 23.0, 27.5, 32.5, and 37.5 as potential public health action points and proposed methods by which countries could make decisions about the definitions of increased risk for their adult populations, although they recommended that the WHO BMI cutoff points (i.e., 25 and 30) should be retained as international classifications.

Adjustment for between-population difference in maturity

Although the WHO Expert Committee suggested that between-population maturity differences should be taken into consideration when interpreting the anthropometric measures on the basis of chronological age, they did not provide specific and practical recommendations for the adjustment of between-population differences in sexual maturation. To our knowledge, few researchers or health professionals have attempted to adjust for maturity differences when reporting their results. Yet sexual maturation is closely associated with growth patterns and nutritional status in children and adolescents. A large and steadily growing body of literature has addressed these issues.

It is well known that there are large between-population differences in the timing and patterns of maturation [50–52, 81]. For example, 11-year-old American girls are likely to be at different maturation stages and to have different growth rates from their counterparts in India. Using US NHANES III data, we found a strong association between maturation and overweight in both girls and boys [89]. Early-maturing girls were twice as likely as average- and late-maturing girls to be classified as overweight. In contrast, early-maturing boys were less likely to be classified as overweight

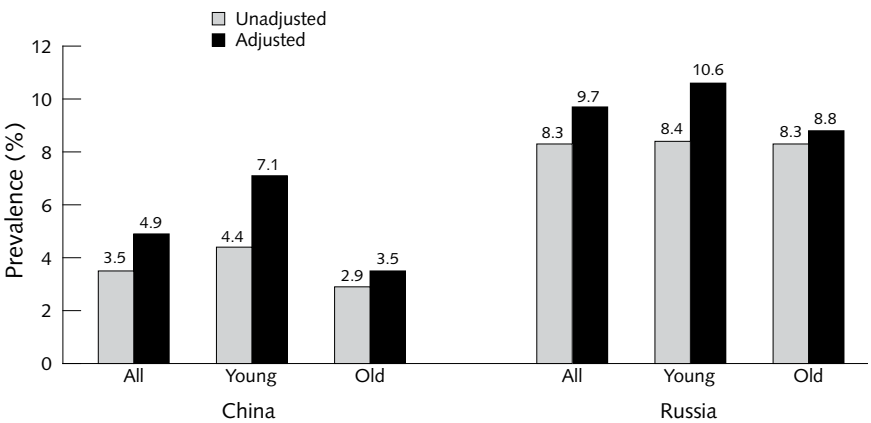

FIG. 2. Unadjusted and maturation-adjusted prevalence of overweight in Chinese and Russian girls according to age group. Young: 10 to 13 years. Old: 14 to 18 years. The adjusted prevalence was calculated by using maturational age–matched body-mass index (BMI) cutoff points (maturational age = chronological age—0.9 for China, and maturational age = chronological age—0.4 for Russia); the unadjusted prevalence was calculated by using chronological age–matched BMI cutoff points. The total sample size was 1,316 for China and 744 for Russia. Adapted from Wang and Adair [90]

(odds ratio, 0.4). Using national survey data for adolescent girls in the United States, China, and Russia, we examined the potential influence of the adjustment of between-population maturation differences on estimates of overweight prevalence [90]. Our analysis suggests that the adjustment could affect the estimates considerably for Chinese and Russian girls (see **fig. 2**). The adjustment increased the prevalence estimate by about one-quarter to one-third (in relative terms) for adolescent girls in China, where children matured later than the reference population, but decreased the estimate in the United States, where children matured earlier. The adjustment had a greater effect in girls around the age of puberty (10 to 13 years) than in older girls (14 to 18 years).

One additional issue that needs to be highlighted is "adiposity rebound," which refers to the second increase in BMI during early childhood [91, 92]. It is of concern that between-population differences in the patterns of adiposity rebound may exist—in particular, between populations in industrialized and developing countries. This may affect the estimate of obesity prevalence for children at around the age of adiposity rebound when the international BMI references based upon data collected in a particular wealthy society are used. Our preliminary analysis shows that differences exist in the timing of adiposity rebound (**fig. 3**), and that there are secular trends toward an earlier age of adiposity rebound in some populations [29]. National representative survey data show that the age of adiposity rebound is around 6 years in France, the Netherlands, and the United Kingdom, 5.5 years in the United States, 5 years in Italy, but 7 years in China. Interestingly, recent data collected from a large sample of 96,104 children in Shanghai, the largest and most prosperous city in China, where the living standard is comparable to that in many industrialized countries, show that the age of adiposity rebound has fallen to around 5 years [29]. The timing of the adiposity rebound corresponds to the degree of BMI centile crossing in individuals or groups [93]. An early rebound indicates centile crossing upwards, whereas a late rebound means that BMI is crossing centiles downwards. This pattern of change in BMI means that the age of adiposity rebound inevitably predicts later BMI, as was first shown by Rolland-Cachera et al. [91]. This association between centile crossing and later changes in BMI applies at all ages, not just at around 5 years, so the adiposity rebound should not be viewed as a critical period for obesity development.

Conclusions

In summary, the NCHS/WHO growth references have played an important role in the past for international use in assessing child and adolescent growth and nutritional status. The US NCHS data collected in the 1960s and 1970s were considered the best available at the time for the development of growth references for international use. However, the NCHS/WHO growth references now suffer from a number of theoretical, methodological, and practical problems. The NCHS/CDC has developed a new set of growth charts to replace these old references in the United States, and WHO has recently developed and recommended the use of a new international reference for infants and preschool children. The global obesity epidemic, which affects both industrialized and developing countries, poses another challenge to the NCHS/WHO references. The history of the development and evolution of the

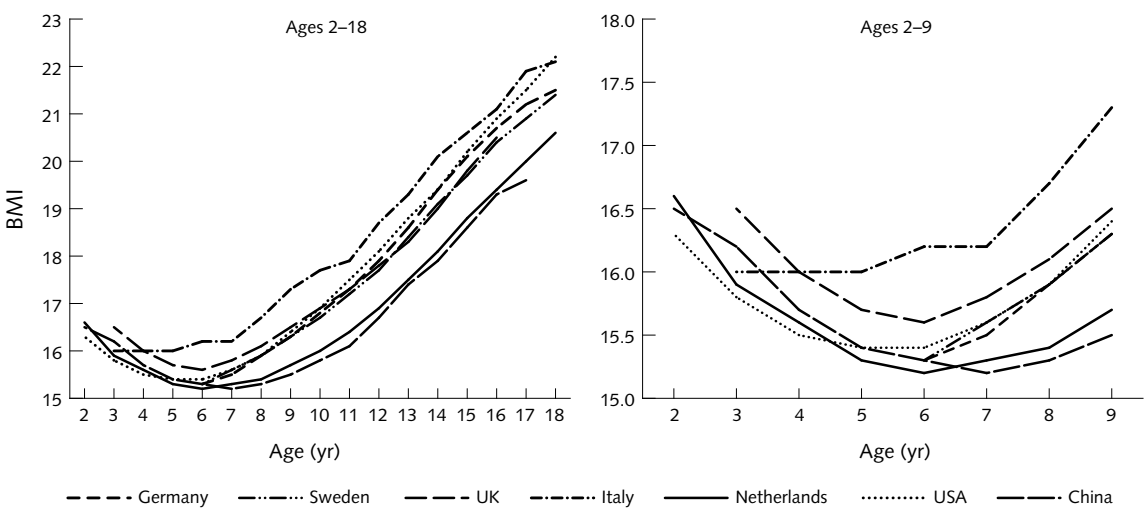

FIG. 3. Median body-mass index (BMI) according to age and adiposity rebound among boys in China, Germany, Italy, the Netherlands, Sweden, the United Kingdom, and the United States. Adapted from Wang [29]

WHO-recommended growth references over the past several decades is a dynamic one. Improvements and changes are often made when adequate new knowledge and better data are available. There is a need for a new international reference to assess the nutritional status and growth of school-aged children and adolescents. A reevaluation of the NCHS/WHO references and the development of a new international growth reference for children and adolescents are therefore two goals of high priority.

Acknowledgments

Research activities related to this work were supported by the Johns Hopkins University and the National Institutes of Health (grant no. 1 R01 DK63383-01). We thank Ms. Diane Reese and Brenda Denzler for their assistance in editing the manuscript, and we are also grateful to the anonymous reviewers.

References

1. de Onis M, Yip R. The WHO growth chart: historical considerations and current scientific issues. Bibl Nutr Dieta 1996;(53):74–89.
2. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser 1995;854:1–452.
3. de Onis M, Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. Am J Clin Nutr 1996;64:650–8.
4. de Onis M, Garza C, Habicht JP. Time for a new growth reference. Pediatrics 1997;100:E8.
5. Garza C, de Onis M, for the WHO Multicentre Growth Referende Group. A new international growth reference for young children. Am J Clin Nutr 1999;70(1 pt 2): 169S–72S.
6. Garza C, de Onis M. Rationale for developing a new international growth reference. Food Nutr Bull 2004;25: S5–14.
7. World Health Organization. An evaluation of infant growth: the use and interpretation of anthropometry in infants. WHO Working Group on Infant Growth. Bull World Health Organ 1995;73:165–74.
8. Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, Mei Z, Curtin LR, Roche AF, Johnson CL. CDC growth charts: United States. Adv Data 2000; (314):1–27.
9. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, Wei R, Curtin LR, Roche AF, Johnson CL. 2000 CDC growth charts for the United States: methods and development. Vital Health Stat 11 2002;(246):1–190.
10. World Health Organization. Use and interpretation of anthropometric indicators of nutritional status. WHO Working Group. Bull World Health Organ 1986;64: 929–41.
11. Hamill PV, Drizd TA. NCHS growth charts. Mon Vital Stat Rep 1976. Report No.: 25 suppl.
12. Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF. NCHS growth curves for children birth–18 years. United States. Vital Health Stat 11 1977;(165)i–iv, 1–74.
13. Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF, Moore WM. Physical growth: National Center for Health Statistics percentiles. Am J Clin Nutr 1979;32:607–29.
14. Dibley MJ, Goldsby JB, Staehling NW, Trowbridge FL. Development of normalized curves for the International Growth Reference: historical and technical considerations. Am J Clin Nutr 1987;46:736–48.
15. Graitcer PL, Gentry EM. Measuring children: one reference for all. Lancet 1981;2(8241):297–9.
16. Sullivan K, Trowbridge F, Gorstein J, Pradilla A. Growth references. Lancet 1991;337:1420–1.
17. World Health Organization. A growth chart for international use in maternal and child healthcare: guidelines for primary health care personnel. Geneva: WHO 1978.
18. Must A, Dallal GE, Dietz WH. Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht^2) and triceps skinfold thickness. Am J Clin Nutr 1991;53:839–46.
19. Owen GM. Measurement, recording, and assessment of skinfold thickness in childhood and adolescence: report of a small meeting. Am J Clin Nutr 1982;35:629–38.
20. Johnson CL, Fulwood R, Abraham S, Bryner JD. Basic data on anthropometric measurements and angular measurements of the hip and knee joints for selected age groups 1–74 years of age. Vital Health Stat 11 1981; (219):1–68.
21. Wang Y, Wang JQ. A comparison of international references for the assessment of child and adolescent overweight and obesity in different populations. Eur J Clin Nutr 2002;56:973–82.
22. de Onis M, Onyango AW. The Centers for Disease Control and Prevention 2000 growth charts and the growth of breastfed infants. Acta Paediatr 2003;92:413–9.
23. Barlow SE, Dietz WH. Obesity evaluation and treatment: Expert Committee recommendations. The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services. Pediatrics 1998;102:E29.
24. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240–3.
25. Himes JH, Dietz WH. Guidelines for overweight in adolescent preventive services: recommendations from an Expert Committee. The Expert Committee on Clinical Guidelines for Overweight in Adolescent Preventive Services. Am J Clin Nutr 1994;59:307–16.
26. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:i–xii, 1–253.
27. Dietz WH, Robinson TN. Use of the body mass index (BMI) as a measure of overweight in children and ado-

lescents. J Pediatr 1998;132:191–3.
28. Prentice AM, Jebb SA. Beyond body mass index. Obes Rev 2001;2:141–7.
29. Wang Y. Epidemiology of childhood obesity—methodological aspects and guidelines: What is new? Int J Obes Relat Metab Disord 2004;28(suppl 3):S21–8.
30. Cole TJ. A method for assessing age-standardized weight-for-height in children seen cross-sectionally. Ann Hum Biol 1979;6:249–68.
31. Cole TJ. Weight/height^p compared to weight/height^2 for assessing adiposity in childhood: influence of age and bone age on p during puberty. Ann Hum Biol 1986;13:433–51.
32. Franklin MF. Comparison of weight and height relations in boys from 4 countries. Am J Clin Nutr 1999;70: 157S–62S.
33. Freedman DS, Thornton JC, Mei Z, Wang J, Dietz WH, Pierson RN Jr, Horlick M. Height and adiposity among children. Obes Res 2004;12:846–53.
34. Reilly JJ, Dorosty AR, Emmett PM; Avon Longitudinal Study of Pregnancy and Childhood Study Team. Identification of the obese child: adequacy of the body mass index for clinical practice and epidemiology. Int J Obes Relat Metab Disord 2000;24:1623–7.
35. Sarria A, Moreno LA, Garcia-Llop LA, Fleta J, Morellon MP, Bueno M. Body mass index, triceps skinfold and waist circumference in screening for adiposity in male children and adolescents. Acta Paediatr 2001;90: 387–92.
36. Wang Y, Rimm EB, Stampfer MJ, Willett WC, Hu FB. Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. Am J Clin Nutr 2005;81:555–63.
37. Koh-Banerjee P, Wang Y, Hu FB, Spiegelman D, Willett WC, Rimm EB. Changes in body weight and body fat distribution as risk factors for clinical diabetes in US men. Am J Epidemiol 2004;159:1150–9.
38. North American Association for the Study of Obesity (NAASO) and the National Heart, Lung, and Blood Institute (NHLBI). The practical guide: identification, evaluation, and treatment of overweight and obesity in adults. 2000. NIH Publication No. 00-4084. Available at: http://www.nhlbi.nih.gov/guidelines/obesity/practgde.htm. Accessesd 4 September 2006.
39. Pelletier D. Theoretical considerations related to cutoff points. Food Nutr Bull 2006;27(suppl):S224–36.
40. Waterlow JC, Buzina R, Keller W, Lane JM, Nichaman MZ, Tanner JM. The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. Bull World Health Organ 1977;55:489–98.
41. Dibley MJ, Staehling N, Nieburg P, Trowbridge FL. Interpretation of Z-score anthropometric indicators derived from the international growth reference. Am J Clin Nutr 1987;46:749–62.
42. Cole TJ, Faith MS, Pietrobelli A, Heo M. What is the best measure of adiposity change in growing children: BMI, BMI%, BMI z-score or BMI centile? Eur J Clin Nutr 2005;59:419–25.
43. Cole TJ. Fitting smoothed centile curves to reference data (with discussion). J R Stat Soc A 1988;151:385–418.
44. Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr 1990;44:45–60.
45. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305–19.
46. Hammer LD, Kraemer HC, Wilson DM, Ritter PL, Dornbusch SM. Standardized percentile curves of body-mass index for children and adolescents. Am J Dis Child 1991;145:259–63.
47. Cole TJ. The secular trend in human physical growth: a biological view. Econ Hum Biol 2003;1:161–8.
48. Deurenberg P, Yap M, van Staveren WA. Body mass index and percent body fat: a meta analysis among different ethnic groups. Int J Obes Relat Metab Disord 1998;22:1164–71.
49. Deurenberg P, Deurenberg-Yap M, Foo LF, Schmidt G, Wang J. Differences in body composition between Singapore Chinese, Beijing Chinese and Dutch children. Eur J Clin Nutr 2003;57:405–9.
50. Eveleth PB, Tanner JM. Worldwide variation in human growth. Cambridge, UK: Cambridge University Press, 1990.
51. Falkner F, Tanner JM. Human growth: a comprehensive treatise. Methodology and ecological, genetic, and nutritional effects on growth. New York: Plenum Press, 1986.
52. Karlberg J. Secular trends in pubertal development. Horm Res 2002;57(suppl 2):19–30.
53. Wells JC, Coward WA, Cole TJ, Davies PS. The contribution of fat and fat-free tissue to body mass index in contemporary children and the reference child. Int J Obes Relat Metab Disord 2002;26:1323–8.
54. Zellner K, Jaeger U, Kromeyer-Hauschild K. Height, weight and BMI of schoolchildren in Jena, Germany—Are the secular changes levelling off? Econ Hum Biol 2004;2:281–94.
55. Habicht JP, Martorell R, Yarbrough C, Malina RM, Klein RE. Height and weight standards for preschool children. How relevant are ethnic differences in growth potential? Lancet 1974;1(7858):611–4.
56. Ashcroft MT, Desai P. Ethnic differences in growth potential of children of African, Indian, Chinese and European origin. Trans R Soc Trop Med Hyg 1977;70:433–8.
57. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. Cross sectional stature and weight reference curves for the UK, 1990. Arch Dis Child 1995;73:17–24.
58. Ge K, Zhai F, Yan H. The dietary and nutritional status of Chinese population: 1992 National Nutrition Survey. Vol 2 (children and adolescents). Beijing: People's Medical Publishing House, 1999.
59. Sempe M, Pedron G, Roy Pernot MP. Auxologie, methodes et sequences. Paris: Theraplix, 1979.
60. Cacciari E, Milani S, Balsamo A, Dammacco F, De Luca F, Chiarelli F, Pasquino AM, Tonini G, Vanelli M. Italian cross-sectional growth charts for height, weight and BMI (6–20 y). Eur J Clin Nutr 2002;56:171–80.
61. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in the Netherlands 1955–1997. Pediatr Res 2000;47:316–23.
62. Hernandez M. Curvas y tabla de crecimiento. Madrid: Garsi, 1988.
63. Lindgren G, Strandell A, Cole T, Healy M, Tanner J.

Swedish population reference standards for height, weight and body mass index attained at 6 to 16 years (girls) or 19 years (boys). Acta Paediatr 1995;84:1019–28.
64. Wikland KA, Luo ZC, Niklasson A, Karlberg J. Swedish population-based longitudinal reference values from birth to 18 years of age for height, weight and head circumference. Acta Paediatr 2002;91:739–54.
65. Goldstein H, Tanner JM. Ecological considerations in the creation and the use of child growth standards. Lancet 1980;1(8168 pt 1):582–5.
66. Beunen GP, Rogol AD, Malina RM. Indicators of biological maturation and secular changes in biological maturation. Food Nutr Bull 2006; 27(suppl):S244–56.
67. Haas JD, Campirano F. Interpopulation variation in height from 7 to 18 years of age. Food Nutr Bull 2006; 27(suppl):S212–23.
68. de Onis M, Dasgupta P, Saha S, Sengupta D, Blossner M. The National Center for Health Statistics reference and the growth of Indian adolescent boys. Am J Clin Nutr 2001;74:248–53.
69. Hakeem R, Shaikh AH, Asar F. Assessment of linear growth of affluent urban Pakistani adolescents according to CDC 2000 references. Ann Hum Biol 2004;31:282–91.
70. Eckhardt CL, Adair LS. Differences in stunting prevalences calculated from two similar growth references may be large and inconsistent in undernourished children. Ann Hum Biol 2002;29:566–78.
71. Moestue H, de Pee S, Hall A, Hye A, Sultana N, Ishtiaque MZ, Huq N, Bloem MW. Conclusions about differences in linear growth between Bangladeshi boys and girls depend on the growth reference used. Eur J Clin Nutr 2004;58:725–31.
72. Lobstein T, Baur L, Uauy R; IASO International Obesity TaskForce. Obesity in children and young people: a crisis in public health. Obes Rev 2004;5(suppl 1):4–104.
73. Wang Y, Must A. Global burden of childhood obesity. In: Kennedy E, Deckelbaum RJ, eds. The nation's nutrition. Washington DC: International Life Sciences Institute (in press).
74. Poskitt EM. Defining childhood obesity: the relative body mass index (BMI). European Childhood Obesity Group. Acta Paediatr 1995;84:961–3.
75. Rolland-Cachera MF, Cole TJ, Sempe M, Tichet J, Rossignol C, Charraud A. Body mass index variations: centiles from birth to 87 years. Eur J Clin Nutr 1991;45:13–21.
76. Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. Arch Dis Child 1995; 73:25–9.
77. Luciano A, Bressan F, Zoppi G. Body mass index reference curves for children aged 3–19 years from Verona, Italy. Eur J Clin Nutr 1997;51:6–10.
78. Group of China Obesity Task Force. Body mass index reference norm for screening overweight and obesity in Chinese children and adolescents. Chin J Epidemiol 2004;25:97–102 (in Chinese).
79. Zimmermann MB, Hess SY, Hurrell RF. A national study of the prevalence of overweight and obesity in 6–12 y-old Swiss children: body mass index, body-weight perceptions and goals. Eur J Clin Nutr 2000;54:568–72.
80. Al-Sendi AM, Shetty P, Musaiger AO. Prevalence of overweight and obesity among Bahraini adolescents: a comparison between three different sets of criteria. Eur J Clin Nutr 2003;57:471–4.
81. Morabia A, Costanza MC. International variability in ages at menarche, first livebirth, and menopause. World Health Organization Collaborative Study of Neoplasia and Steroid Contraceptives. Am J Epidemiol 1998;148:1195–205.
82. Power C, Lake JK, Cole TJ. Measurement and long-term health risks of child and adolescent fatness. Int J Obes Relat Metab Disord 1997;21:507–26.
83. Serdula MK, Ivery D, Coates RJ, Freedman DS, Williamson DF, Byers T. Do obese children become obese adults? A review of the literature. Prev Med 1993;22:167–77.
84. Wang Y, Ge K, Popkin BM. Tracking of body mass index from childhood to adolescence: a 6-y follow-up study in China. Am J Clin Nutr 2000;72:1018–24.
85. Mo-suwan L, Tongkumchum P, Puetpaiboon A. Determinants of overweight tracking from childhood to adolescence: a 5 y follow-up study of Hat Yai schoolchildren. Int J Obes Relat Metab Disord 2000;24:1642–7.
86. Chinn S, Rona RJ. International definitions of overweight and obesity for children: a lasting solution? Ann Hum Biol 2002;29:306–13.
87. Reilly JJ. Assessment of childhood obesity: national reference data or international approach? Obes Res 2002;10:838–40.
88. World Health Organization. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363(9403):157–63.
89. Wang Y. Is obesity associated with early sexual maturation? A comparison of the association in American boys versus girls. Pediatrics 2002;110:903–10.
90. Wang Y, Adair L. How does maturity adjustment influence the estimates of overweight prevalence in adolescents from different countries using an international reference? Int J Obes Relat Metab Disord 2001;25: 550–8.
91. Rolland-Cachera MF, Deheeger M, Bellisle F, Sempe M, Guilloud-Bataille M, Patois E. Adiposity rebound in children: a simple indicator for predicting obesity. Am J Clin Nutr 1984;39:129–35.
92. Whitaker RC, Pepe MS, Wright JA, Seidel KD, Dietz WH. Early adiposity rebound and the risk of adult obesity. Pediatrics 1998;101:E5.
93. Cole TJ. Children grow and horses race: Is the adiposity rebound a critical period for later obesity? BMC Pediatr 2004;4:6.

Cross-sectional growth references and implications for the development of an international growth standard for school-aged children and adolescents

Jacob C. Seidell, Colleen M. Doak, Jeroen S. L. de Munter, Lothar D. J. Kuijper, and Cor Zonneveld

Abstract

Normative data are needed to create a reference that indicates optimal development of weight in relation to height and age, particularly in the face of the unfolding obesity epidemic. The body-mass index (BMI) has some serious limitations: it is a relatively poor predictor of current and future fatness. Currently, however, there are few available alternatives, with the possible exception of waist circumference or skinfolds. The use of cross-sectional references to construct a BMI-reference curve is problematic when there are period and cohort effects. Ideally, a reference would be based on longitudinal data in populations with little underweight, overweight, and obesity.

In the meantime cross-sectional data in appropriate populations could be used to construct BMI percentiles linking BMI values at age 5 to those at age 18 (or 21) that would correspond with adult BMI values reflecting optimal health (e.g., that would correspond to adult BMI values between 21 and 23 kg/m²).

Key words: Growth reference, school-aged children, longitudinal data

Introduction and rationale

Background

In 2003, a meeting in Rome brought together representatives from the Department of Nutrition for Health and Development of the World Health Organization (WHO), the United Nations University Food and Nutrition Program (UNU-FNP), and the Food and Agriculture Organization (FAO) to consider the feasibility and appropriateness of developing a single international growth reference or standard for school-aged preadolescents and adolescents. In particular, the attendees outlined a process for evaluating the potential content and appropriateness of an international growth standard presenting children's optimum growth, rather than a growth reference describing the current growth status of a particular population, some of whom may not be growing optimally.

The purpose of this chapter is to describe the uses of cross-sectional growth references and their implications for the development of an international growth reference or standard.

References for specific growth measurements or indicators can be developed for different purposes [1]:

» Identification of individuals or populations at risk for disease;
» Selection of individuals or populations for interventions;
» Evaluation of the impact of interventions;
» Excluding individuals or populations from interventions (i.e., those predicted to be not at risk);
» Achieving normative standards.

Ideally, references are suitable for all of these purposes, at both the individual and the population level (e.g., to estimate prevalences, estimate time trends, and make comparisons between populations).

Criteria for an international growth reference

Prescriptive approaches for the development of growth standards should describe the height and weight of populations in which undernutrition ("malnutrition") and overnutrition are virtually absent and the prevalence of conditions related to suboptimal nutrition or diseases that affect optimal growth is low. This means that, if such data exist, the data should be taken from populations with an exceptionally low prevalence of stunting, infectious diseases, and noncommunicable chronic diseases, such as cardiovascular disease, type 2 diabetes, musculoskeletal disorders, and cancer.

The authors are affiliated with the Institute for Health Sciences, Vrije Universiteit Amsterdam, Amsterdam.

Please direct queries to the corresponding author: Jacob C. Seidell, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; e-mail: jaap.seidell@falw.vu.nl.

TABLE 1. Common terms for height- and weight-based anthropometric indicators

Anthropometric indicator	Terms describing outcomes	Terms describing process	Explanation
Low height-for-age	Shortness Stunted	Stunting (gaining insufficient height relative to age)	Descriptive Implies long-term malnutrition and poor health
Low weight-for height	Thinness Wasted	Wasting (gaining insufficient weight relative to height or losing weight)	Descriptive Implies recent or continuing current severe weight loss
High weight-for-height or high BMI	Heaviness Overweight	Gaining excess weight relative to height, or gaining insufficient height relative to weight	Descriptive Implies obesity
Low weight-for-age	Lightness Underweight	Gaining insufficient weight relative to age, or losing weight	Descriptive Implies stunting and/or wasting
High weight-for-age	Heaviness Overweight	Gaining excess weight relative to age	Descriptive Implies overweight as a result of obesity

BMI, body-mass index
Source: [1].

Table 1 shows the common terms used to describe deviations from optimal growth on the basis of height-for-age, weight-for-height, weight-for-age, and body-mass index (BMI), based on the 1995 WHO Expert Committee report on physical status [1]. This table has no category deviating from optimal health that describes high height-for-age (tallness). Such a category might be useful, because during the last 30 years, several researchers have found a negative correlation between greater height and longevity on the basis of relatively homogeneous deceased population samples [2]. Studies suggest that people with shorter, smaller bodies have lower death rates and fewer diet-related chronic diseases, especially after middle age. Shorter people also appear to have longer average lifespans. It has been suggested that the differences in longevity between the sexes is due to their height difference, because men on average are 8.0% taller than women and have a 7.9% lower life expectancy at birth. Animal experiments also show that smaller animals within the same species generally live longer. The relation between height and health has become more important in recent years, because rapid developments in genetic engineering may offer parents the opportunity to increase the heights of their children. Increasing the proportion of taller and heavier people over the generations without careful consideration of the impact of this increase may have a negative effect on the health of populations. The increased risk of several important types of cancer, such as colon and breast cancer, in taller people and populations may be a particular concern [3].

To the best of our knowledge, there is no objective inventory of countries or populations in which health is systematically assessed that allows the identification of a population with optimal health. In fact, WHO commonly divides the world into three regions with regard to health:

» Developing countries with high mortality (the poorest nations, with high prevalences of infectious diseases, undernutrition, and stunting);
» Developing countries with low mortality (e.g., countries undergoing economic transition and experiencing the double or triple burden of disease, which implies the coexistence of a high prevalence of diseases of poverty, emerging chronic diseases, and injuries);
» Developed countries with a high prevalence of non-communicable diseases.

This means that, currently, there are no regions with optimal health (i.e., low prevalence of diseases of poverty, chronic diseases, and injuries).

It is likely that, when undergoing an economic transition, populations have experienced a rapidly declining incidence of communicable diseases and diseases related to poverty before the onset of non-communicable diseases. We propose that somewhere between the 1950s and the 1970s, the developed countries (or subpopulations of countries) came close to the description of optimal health some time after the introduction of major public health measures, such as improved sanitation, increased food availability, and statewide immunization programs, but before the

epidemic increases in cardiovascular disease, type 2 diabetes, and cancer.

Current international references

De Onis et al. [4] published the results of a survey across United Nations member states on the uses of growth references. A questionnaire was sent to Ministries of Health in 202 countries requesting information about growth charts for various uses. Responses were received from 178 countries, 154 of which included growth charts. Although the study focused on children from birth to 6 years of age, it is clear that only a small fraction (25 of the responding countries) used national growth charts, and the rest used the references of the US Centers for Disease Control and Prevention/National Center for Health Statistics (CDC/NCHS), which are described in more detail in the following section. More affluent countries are more likely to use national reference charts, and it may be that, as is the case for the Netherlands, combinations of national references and International Obesity Task Force (IOTF) cutoff points for overweight are used simultaneously. Globally, however, the CDC charts are the most widely used.

The CDC growth references

The most commonly used growth references are those of the US Centers for Disease Control and Prevention (CDC).* The growth charts consist of a series of percentile curves that illustrate the distribution of selected body measurements in US children. The 1977 growth charts were developed by CDC/NCHS as a clinical tool for health professionals to determine if the growth of a child is adequate. The 1977 charts were adopted by WHO for international use. When the 1977 CDC/NCHS growth charts were first developed, CDC/NCHS recommended that they be revised periodically as necessary. With more recent and comprehensive national data now available, along with improved statistical procedures, the 1977 growth charts were revised and updated to make them a more valuable clinical tool for health professionals. The 2000 CDC/NCHS growth charts represent the revised version of the 1977 CDC/NCHS growth charts. Most of the data used to construct these charts come from the National Health and Nutrition Examination Survey (NHANES), which has periodically collected information on height, weight, and other health-related features from the American population since the early 1960s.

Growth charts are not intended to be used as a sole diagnostic instrument. Instead, growth charts are tools that contribute to forming an overall clinical impression of the child being measured. The revised growth charts provide an improved tool for evaluating the growth of children in clinical and research settings.

The 2000 CDC growth charts and the new BMI-for-age charts

The addition of the BMI charts is probably the single most significant new feature of the revised growth charts [5]. These BMI-for-age charts were created for use in place of the 1977 weight-for-stature charts. BMI (the weight in kilograms divided by the square of the height in meters) is used to judge whether an individual's weight is appropriate for his or her height. BMI is the most commonly used measure to determine if adults are overweight or obese and is also the recommended measure to determine if children are overweight. The new BMI growth charts can be used clinically for children beginning at 2 years of age, when an accurate height can be obtained. In recent years, BMI has received increased attention for pediatric use. In 1994, an expert committee charged with developing guidelines for overweight in adolescent preventive services (ages 11 to 21 years) recommended that BMI be used routinely to screen for overweight adolescents. In addition, in 1997 an expert committee on the assessment and treatment of childhood obesity concluded that the BMI curves from the revised growth charts should be used to screen for overweight children, aged 2 years and older. BMI can also be used to characterize underweight, although no expert guidelines exist for the classification of underweight on the basis of BMI.

Each of the CDC BMI-for-age sex-specific charts contains a series of curved lines indicating specific percentiles. **Table 2** shows the established cutpoints used by healthcare professionals to identify underweight and overweight in children.

Chinn [6] has argued that because the current CDC reference distribution is composed of data from five surveys, it does not represent the US population at any one time. Also, the rationale for using the percentiles to diagnose underweight is described poorly.

TABLE 2. Cut-points for body-mass index (BMI) to identify overweight and underweight in children established by the Centers of Disease Control, 2000

Underweight	BMI-for-age < 5th percentile
Normal	BMI-for-age 5th percentile to < 85th percentile
At risk of overweight	BMI-for-age 85th percentile to < 95th percentile
Overweight	BMI-for-age ≥ 95th percentile

* The CDC growth charts are available at: http://www.cdc.gov/nccdphp/dnpa/growthcharts/training/modules/module2/text/page5i.htm.

The International Obesity Task Force (IOTF) classification of overweight in children and adolescents

Growth standards aim to describe growth in the absence of underweight, stunting, or overweight. Many attempts in the past have focused on identification of children and adolescents at the extremes of the distributions of weight-for-height or BMI. One such example is the IOTF proposal for an international definition of overweight and obesity using BMI [7].

For each of six surveys (from six different countries: Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States), centile curves for BMI were drawn so that at the age of 18 years, the 85th and 95th centiles passed through the widely used cutoff points of 25 and 30 for adult overweight and obesity, respectively. The resulting curves were averaged to provide age- and sex-specific cutoff points from 2 to 18 years [7].

Ideally, the definitions of overweight and obesity should be based on the risk of morbidity, but there is currently insufficient information about children to allow such classification. Internationally accepted definitions are desirable to allow for comparisons between populations and to evaluate time trends. There is no international consensus on which pediatric cutoff points should be used. Chinn [6] has cited five objections to the use of the IOTF cutoff points:

» Z-scores of BMI are required;
» No cutoff points for underweight and severe obesity are designated;
» The IOTF may under- or overestimate overweight in comparison to national definitions;
» The IOTF does not include children under the age of 2 years;
» If comparison with an earlier study for which the raw data are not available is required, then there will be no alternative but to use the earlier study's definition. But this does not preclude the additional use of the IOTF definition where data are available.

Other critics of the IOTF cutoff points include Reilly [8, 9], who argued that the sensitivity of the IOTF cutoff points was low to detect obesity defined by using bioelectrical impedance and high when UK cutoff points are used. The inherent limitations of BMI to reflect body composition, however, cannot be addressed by modifying cutoff points.

Flegal et al. [5] demonstrated that the CDC definition of obesity results in higher prevalences of obesity than the IOTF definition, but this can be explained by the fact that the CDC uses the 95th percentile and the IOTF uses the 96th or 97th percentile, corresponding to a BMI of 30 kg/m^2 at the age of 18.

Although the IOTF cutoff points have been criticized, the idea that BMI criteria by age for children and adolescents should, at ages reaching adulthood, correspond to the BMI cutpoints used in adults is appealing. If not, an individual may be considered obese according to criteria for adolescents at age 18 and 19 but not according to criteria for adults and this may lead to confusion for both the individual and their healthcare providers.

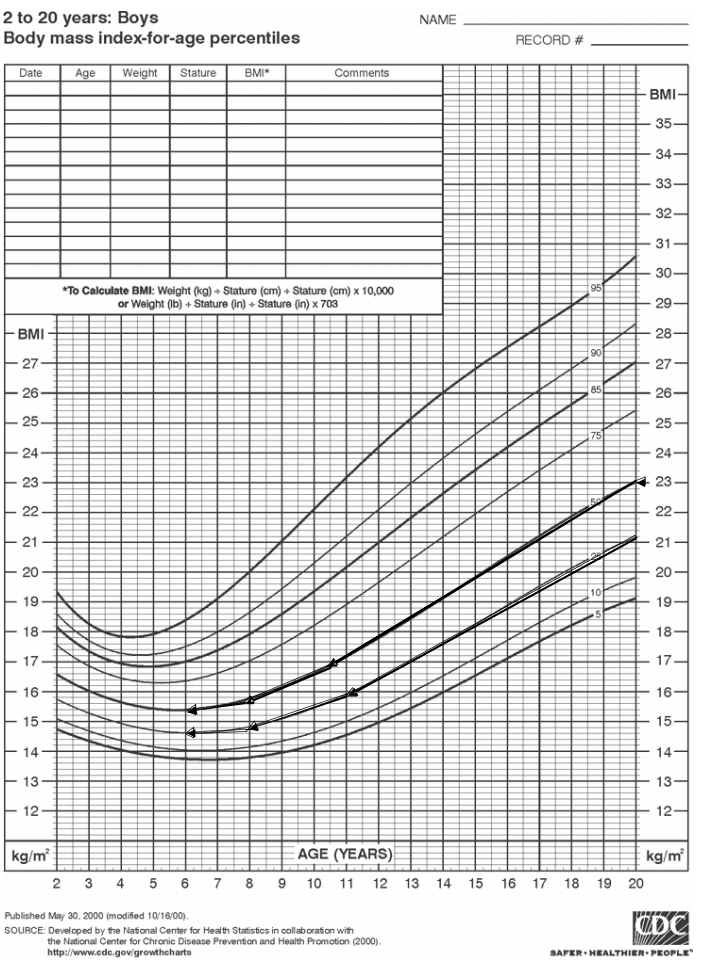

FIG. 1. Centiles of body-mass index (BMI)-for-age in the US population (CDC). The centiles are marked that correspond to the adult range of BMI between 21 and 23, which is defined as optimal for adults by WHO [1].

Using the IOTF rationale for establishing a definition of optimal growth

Several WHO reports have suggested optimal BMIs for adult populations. The report on "Diet, nutrition and the prevention of chronic diseases" [10] states (page 69) that "to achieve optimum health the median population should be in the range of 21–23 kg/m^2, while the goal for individuals should be to maintain BMI in the range of 18.5–24.9 kg/m^2."

In order to establish whether or not BMI develops optimally, it would be possible to have BMI "isobars" that correspond with BMIs at 21, 22, and 23 for adults. It may be advisable to use BMI at 19 or 20 years of age rather than 18 years as the adult reference [11] and to use national centiles rather than Cole's centiles based on a mixture of countries.

Ideally, these "centiles" would also correspond to the new WHO growth standards for children aged 5 years or younger. **Figure 1** shows an example using the CDC centiles for boys. A BMI of 23 at the age of 20 corresponds to the 50th percentile, and a BMI of 21 to the 25th percentile. This would imply that optimal growth in the United States therefore should be maintained between the 25th and 50th BMI percentiles.

Methodological limitations of cross-sectional growth references

One assumption of using BMI-for-age centiles is that only age effects are described and that there are no period or cohort effects. Thus, children are expected to follow these centile lines as they age. If there are major cohort effects, for instance, an increase in BMI over time, as seen in many populations across the world, longitudinal centile lines will deviate from the cross-sectional centile lines.

An example of this is seen in **fig. 2**, derived from a mixed longitudinal study of adults in the Netherlands [12]. The "longitudinal" curve is an estimate based on linking the longitudinal 10-year follow up. Ideally, however, a reference population should be based on a population with little or no secular trends in BMI for the relevant age groups.

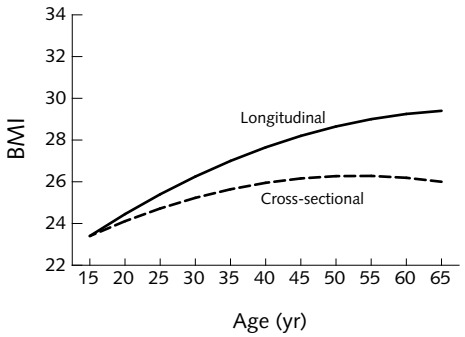

FIG. 2. Relationship of body-mass index (BMI) and age based on cross-sectional data and on a projection of longitudinal 10-year follow-up data in a mixed longitudinal design from the Netherlands. After Nooyens et al. [12]

Consequently, BMI may be a poor determinant of adult BMI or fatness. For instance, the Amsterdam Longitudinal Study of Growth and Health monitored adolescents from the age of 12 onwards in 1970. In these children, at the ages of 12 to 16, no boys and only very few girls met the criteria of overweight according to the IOTF cutoff points, yet a considerable proportion of these children ended up as overweight adults (BMI > 25) [13]. In addition, the BMI has poor validity for the measurement of body composition in adolescents (particularly in boys) [13, 14] and as an outcome of lifestyle interventions, particularly when these involve an increase in physical activity [15].

A comparison of BMI development in children and adolescents in different national datasets

Main objectives

Our objective in this paper is to identify samples of adolescents from different populations who have minimal undernutrition and minimal overweight on the basis of socioeconomic status. We also considered whether such prescriptive samples could provide the basis for a possible reference population by comparing it with survey data from the Netherlands [16], and whether the prescriptive sample could be used to reflect optimal BMI for all children. Afterwards, we calculated the relationships of BMI with height and age. This was done to better understand the relationship between height and BMI and to explore whether differences in BMI between Asian and non-Asian populations are driven by height differences. Finally, we explored whether scaling for height differences improved BMI comparisons.

Methods

Subjects

This comparison used national survey data from seven different countries. These large representative datasets are from Brazil, the United States, China, Indonesia, Kyrgyzstan, Russia, and the Netherlands. All analyses were performed on adolescents aged 10 to 18 years. Pregnant adolescents were excluded. Other exclusion criteria were missing height, weight, or socioeconomic status determinants. The sample sizes are listed below.

Brazil: Pesquisa Nacional sobre Saude e Nutricao (PSNS, 1989). The survey was conducted by the Insti-

tuto Brasileiro de Geografia e Estatistica (IBGE), the federal agency in charge of national statistics. The sample included 11,500 adolescents; of these, 396 were excluded from the analyses.

China: Health and Nutrition Survey (CNHS, 1993). The CHNS is a large national longitudinal survey covering eight provinces. The provinces were chosen to reflect the variability in geography and economic development of China. The sample included 2,009 adolescents; of these, 379 were excluded from the analyses.

Indonesia: Family Life Survey (IFLS, 1993). The 1993 IFLS was the first wave of a longitudinal survey conducted in 321 communities and 13 provinces by the RAND Corporation in collaboration with Lembaga Demografi, University of Indonesia. The survey is representative of 83% of the population of Indonesia. The survey included 3,520 adolescents; of these, 288 were excluded from the analyses.

Kyrgyz Republic: Multipurpose Poverty Survey (KMPS, 1993). The survey was conducted under the direction of researchers from the University of North Carolina at Chapel Hill, Paragon Research International, and the Institute of Sociology of the Russian Academy of Sciences. The survey was nationally representative. We included 1,076 adolescents in the analyses.

Russia: Longitudinal Monitoring Survey (RLMS, 1996). The survey was based on round 7 of the RLMS. The survey included 1,942 adolescents; of these, 138 were excluded from the analyses.

United States: Third National Health and Nutrition Examination Survey (1991). NHANES III was conducted from October 1988 through October 1994 in two phases, each of which included a national probability sample. The survey design used stratified, multistage probability analyses. In 1991, data were available from 3,600 adolescents; of these, 1,063 were excluded from the analyses.

Netherlands: Third Growth Study (1980). The third growth study was nationally representative. We included 16,524 adolescents in the analyses.

Of the survey data, we used the sex, height, age, and income or education variables. All data were analyzed by SAS version 8.02. We used Epi Info's Nutstat to calculate the height-for-age z-scores and the BMI-for-age z-scores.

Definitions of undernutrition and overweight

Undernutrition (stunting) was defined as a height-for-age z-score (HAZ) < –2 of the 1977 CDC/WHO reference standards. The cutoff point for overweight was the 85th BMI percentile from the 2000 CDC/NCHS growth charts. To identify the income group with the least malnutrition, we determined the prevalence of undernutrition (HAZ < 2) and overweight (BMI > 85th percentile) in five different socioeconomic status groups.

Comparison of socioeconomic status

In Brazil, Indonesia, China, the Kyrgyz Republic, and Russia, we used per capita household expenditures as an indicator of socioeconomic status. In the US NHANES III, a proportional scale was used for income. In the Dutch dataset, we used level of education as our socioeconomic status variable. Five equal groups were created per country according to total household expenditures, household income, or level of education. Thereafter, we examined the prevalence of stunting and overweight in each of these groups. Because only the Dutch dataset used level of education, we shall refer to the groups as "income" groups.

Minimal stunting and minimal overweight

In countries with a high prevalence of undernutrition, the samples for our reference were drawn only from high- and middle-income groups, depending on the association between income and obesity. If the prevalence of obesity increased rapidly in the highest-income group, samples were drawn only from the middle-income groups.

Odds ratios were used to determine the inclusion of children from each income group, selecting only those income groups that reflect optimal health and balancing stunting and overweight. The results were compared across populations and to the Dutch dataset by one-way ANOVA.

Height–age relationship and BMI

Height and age are both associated with changes in BMI. We compared the relationship of height and age with BMI across countries. This was done to better understand the relationship between height and BMI and to explore whether differences in BMI between Asian and non-Asian populations are driven by height differences. For comparative purposes, we have taken the top 80th income percentile and above for all countries. Our main reason for selecting this upper percentile is that height-related undernutrition (i.e., stunting) is not very common in this income group. When, for any reason, the upper-income group had a higher prevalence of stunting than the middle-income group, we used the middle-income group in these analyses.

We performed multiple linear regression analysis, defined in terms of the following formula, to check for an interaction between height and age:

$$\mathrm{BMI} = \beta_0 + \beta_1 * \mathrm{age} + \beta_2 * \mathrm{height} + \beta_3 * (\mathrm{height}*\mathrm{age})$$

Second, we performed stepwise multiple linear regression analysis.

We used height and age as our independent variables and BMI as the dependent variable:

$$BMI = \beta_0 + \beta_1 * age + \beta_2 * height$$

This regression model, bearing BMI as the dependent variable, explained how height and age are associated with changes in BMI. If height or age is not significantly associated with BMI, the stepwise regression will exclude the nonsignificant variable.

Adjustment for height

As stated in the Introduction and Rationale, we are interested in comparing data from children who grow and mature at different rates in different countries. One way to measure a child's relative height growth is to compare the current height with the final adult height in that population. By using this method, we get a proportional height compared with the final adult height. We used height as a proxy for maturation.

We calculated the median final height for boys and girls in all populations. Data from the Netherlands and China were used as examples. We used the Dutch population in 1980 as our reference population, because at that time there was virtually no undernutrition and a low prevalence of overweight (5% to 7% according to the IOTF cutoff points). If the plots between the countries differed, we transformed the age variable to find any maturation differences.

Results

Minimal stunting and overweight

The prevalence of overweight in Indonesia and China was low among all income groups (**table 3**). In Indonesia and China, the prevalence of stunting was lower in higher-income groups (odds ratios for income group 5 versus 1, 0.31 for Indonesia and 0.32 for China).

The prevalence of overweight in Russia was lower in the higher-income groups. We found the highest prevalence of overweight in the lowest income group (odds ratio, 0.82 for income groups 3–5 versus 1–2). The prevalence of stunting was almost equal in all income groups, except for the lowest group (odds ratio, 0.62 for income groups 3–5 versus 1–2). Kyrgyzstan, which is geographically located between the Asian countries and Russia, showed an "Asian" trend in stunting prevalence (odds ratio, 0.6 for income groups 3–4 versus 1–2). In the Kyrgyz Republic, the prevalence of overweight was highest in the lowest-income groups (odds ratio, 0.75 for income groups 3–4 versus 1–2). The prevalence of overweight in the Kyrgyz Republic showed a high-low-high pattern over the income groups.

In Brazil, high-income groups had a low prevalence of stunting (odds ratio, 0.19 for income group 5 versus 1) and a relatively high prevalence of overweight (odds ratio, 3.54 for income group 5 versus 1). In the United States, the prevalence of stunting was below 5% in all income groups, and the prevalence of overweight was lower in the upper-income groups than in the lower-income groups (odds ratio, 0.66 for income group 5 versus 1).

TABLE 3. Prevalence of overweight and stunting as percentage of the population for categories (country-specific quintiles) of income (1 = lowest; 5 = highest) in different countries

		Income group				
Country	Condition	I	II	III	IV	V
Indonesia	Overweight	4	4	4	4	5
	Stunting	60	56	54	42	33
China	Overweight	6	4	3	6	8
	Stunting	34	31	22	20	14
Kyrgyz Republic	Overweight	20	18	15	15	18
	Stunting	35	32	25	24	25
Russia	Overweight	13	10	11	9	9
	Stunting	10	6	4	5	6
Brazil	Overweight	4	6	7	7	14
	Stunting	33	25	18	13	7
United States	Overweight	32	32	27	30	24
	Stunting	4	4	3	2	1
Netherlands	Overweight	6	7	6	6	3
	Stunting	1	1	1	1	1

The US and Brazilian data both showed a high prevalence of overweight in the upper-income groups. Therefore, we split the upper-income group (> 80th percentile) for Brazil and the United States (**table 4**). In the United States, the highest income group had the lowest prevalence of overweight. In Brazil, the exact opposite was observed, with the highest-income group having the highest prevalence of overweight.

Height–age relationship and BMI

We found an interaction between the effects of height and age on BMI in girls from Brazil and the Netherlands (**fig. 3**). In these girls, the effect of height on BMI changes with increasing age.

Multicollinearity

From a biological point of view, both height and age may serve as a proxy for maturation. They both measure the approximate timing of maturation. When height was plotted against BMI we observed a relatively high correlation coefficient in all countries (~ 0.7 < *r* < 0.8).

Table 5 shows the output from the linear regression model. Height predicted significant changes in BMI for boys in six out of seven countries and in three out of seven countries for girls.

Adjustment for height

The median adult height for Dutch and Chinese adolescents was 182 cm (Netherlands) and 167 cm (China) in boys and 168 cm (Netherlands) and 157 cm (China) in girls. The difference in height between Dutch and Chinese boys and girls was 14.5 and 11.5 cm, respectively. The adult height values were used to calculate the scaled height. **Figure 3A** shows the graph of Dutch and Chinese boys on the transformed height axis. The Dutch and Chinese girls showed a different pattern (**fig. 3B**). Multiplying the Chinese age by a factor of 0.97, based on visual comparison, resulted in a growth curve comparable with the Dutch data (**fig. 3C**). Application of this procedure to data from other countries gave similar results.

TABLE 4. Prevalence of overweight among Brazilian and US adolescents from upper-income families

Adolescents	Prevalence of overweight (%)		Odds ratio
	Income percentile 80–90	Income percentile 90+	
Brazil			
Male	10	16	1.82
Female	11	20	1.91
Total	10	17	1.89
USA			
Male	24	19	0.76
Female	29	20	0.61
Total	27	20	0.68

Discussion

We used WHO definitions for overweight and stunting, which resulted in a high prevalence of stunting for the Asian countries. It is questionable, given American standards, how undernourished these Asian children truly are. Although a much lower prevalence is seen in higher-income groups, the absolute prevalence of

TABLE 5. Multiple linear regression results according to country

Country	Gender	Linear regression, BMI dependent	*p*, age	*p*, height	SES %ile
Indonesia	Boys	3.723 + 0.279*age + 0.258*height	< 0.001	< 0.001	80+
	Girls	2.052 + 0.341*age + 0.249*height	< 0.001	< 0.001	80+
China	Boys	11.645 + 0.172*height		0.032	80+
	Girls	11.290 + 0.380*age	< 0.001		80+
Russia	Boys	9.786 + 0.520*age	< 0.001		80+
	Girls	0.048 + 0.198*age + 0.309*height	0.033	0.001	60–80
Kyrgyz Republic	Boys	22.803 + 0.455*age + –0.515*height	< 0.001	< 0.001	40–80
	Girls	21.634 + 0.526*age + –0.565*height	< 0.001	< 0.001	40–80
Brazil	Boys	2.688 + 0.473*height		< 0.001	80+
	Girls	Interaction between age and height			80+
USA	Boys	0.715 + 0.417*height		< 0.001	80+
	Girls	10.263 + 0.360*age	< 0.001		80+
Netherlands	Boys	5.824 + 0.306*age + 0.274*height	< 0.001	< 0.001	All
	Girls	Interaction between age and height			

SES, socioeconomic status

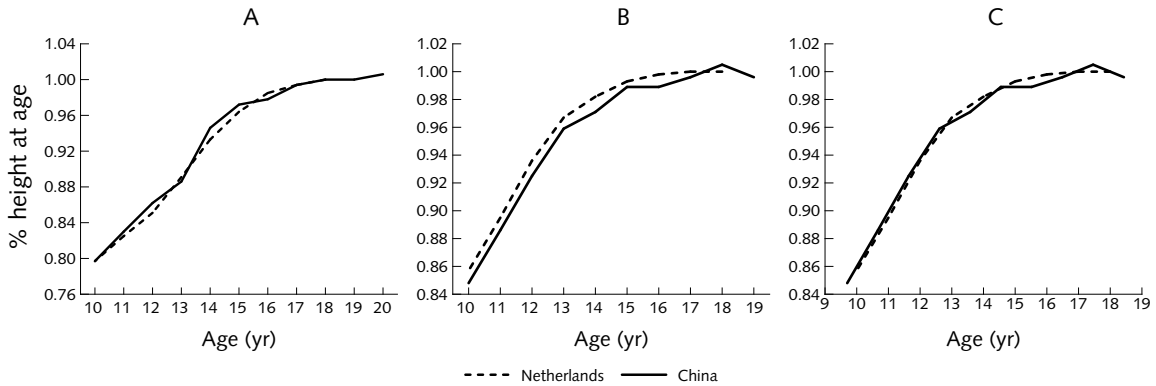

FIG. 3. A. Data from boys from the Netherlands and China plotted on a scaled height axis. B. Data from girls from the Netherlands and China plotted on a scaled height axis. C. Data from Chinese girls scaled to data from girls from the Netherlands

stunting remains high. In our analyses, we have shown that it is possible to select prescriptive samples from each dataset that can be used to build an international growth reference. However, the Dutch dataset shows a lower prevalence of stunting and overweight in all income groups. Given the large sample size in the Dutch dataset, it may be preferable to use it as the prescriptive international standard.

The US NHANES III data from 1991 show a very high prevalence of overweight in adolescents, and therefore these data were excluded from the 2000 CDC reference. Even selecting a "relative" prescriptive subsample from the highest-income group resulted in a prevalence of overweight of greater than 20%. Thus, it is advisable to use only US data obtained prior to the obesity epidemic.

Maturation differences between populations play a major part in the age- and sex-dependent distribution of BMI. The transformation process using adolescent height partially adjusts for these maturation differences. The multiple linear regressions show that the correlation between height and BMI is significant. Transforming actual height to achieved adult height in the population is a way to adjust for height differences among populations. We are still developing this promising method and modeling optimal international BMI cutoff values using this procedure.

It is important to keep in mind that in order to develop international BMI cutoff points, strong evidence-based research on health outcomes related to overweight in children and adolescents is needed. Without this information, the interpretation of BMI thresholds will be more or less arbitrary.

Conclusions

There is no single international growth pattern for adolescents. Populations from different countries show different growth patterns. It is possible to select subpopulations within countries that show a relatively low prevalence of overweight and undernutrition. In this paper, subpopulations were selected on the basis of a measure of socioeconomic status (per capita household income, expenditures, or level of education). The distributions of BMI values of these prescriptive subpopulations showed a more uniform pattern and may be used to further develop prescriptive references for BMI.

When the BMI distribution of the Dutch dataset is compared with those of the selected subpopulations, the Dutch show the lowest prevalence of overweight and stunting in all income groups. These data show that the Dutch population may serve as a better international reference than the prescriptive subpopulations within diverse countries.

It is possible to transform height and to compare the BMI distributions of populations with height differences. This method could be extended and used to calculate national BMI reference charts from one single international reference based on a prescriptive population with minimal stunting and overweight.

This approach to determining optimal BMI cutoff points differs from those used for the current sets of BMI cutoff points for children. Future development of an international prescriptive BMI reference can be used to reflect optimal BMI for all children of the same age, sex, and height and may be used to target further international research and interventions for overweight children.

The following conclusions can be drawn from this study:

» We need normative data for the development of a BMI reference, particularly in the face of the unfolding obesity epidemic;
» BMI has some limitations (it is a relatively poor predictor of current and future fatness), but there are few available alternatives (except, possibly, waist circumference or skinfolds);
» The use of cross-sectional references is problematic

when there are period and cohort effects;

» "Isobars" of BMI percentiles linking BMI values at age 5 to those at age 18 (or 21) would correspond with adult BMI values that, in terms of health risks, could be interpreted differently in diverse populations.

Acknowledgments

No outside funding or support was provided for this study. We acknowledge Roede en Van Wieringen for giving us permission to use a selection from the Dutch dataset, and we would like to thank Stef van Buuren from the Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO) for his cooperation in this process.

References

1. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser 1995;854:1–452.
2. Samaras TT, Elrick H, Storms LH. Is height related to longevity? Life Sci 2003;72:1781–1802.
3. Batty GD, Shipley MJ, Langenberg C, Marmot MG, Davey Smith G. Adult height in relation to mortality from 14 cancer sites in men in London (UK): evidence from the original Whitehall study. Ann Oncol 2006; 17:157–66.
4. de Onis M, Wijnhoven TM, Onyango AW. Worldwide practices in child growth monitoring. J Pediatr 2004; 144:461–5.
5. Flegal KM, Ogden CL, Kuczmarski RL, Johnson CL. Prevalence of overweight in US children: comparison of US growth charts from the Centers for Disease Control and prevention with other reference values for body mass index. Am J Clin Nutr 2001;73:1086–93.
6. Chinn S. Definitions of childhood obesity: current practice. Eur J Clin Nutr 2006;60:1189-94.
7. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240–3.
8. Reilly JJ. Assessment of childhood obesity: national reference data or international approach? Obes Res 2002; 10:838–40.
9. Reilly JJ, Dorosty AR, Emmett PM; Avon Longitudinal Study of Pregnancy and Childhood Study Team. Identification of the obese child: adequacy of the body mass index for clinical practice and epidemiology. Int J Obes Relat Metab Disord 2000;24:1623–7.
10. Diet, nutrition and the prevention of chronic diseases. World Health Organ. Tech Rep Ser 2003;916:i–viii, 1–149.
11. Chinn S, Rona RJ. International definitions of overweight and obesity for children: a lasting solution? Ann Hum Biol 2002;29:306–13.
12. Nooyens ACJ, Koppes LLJ, Visscher TLS, Twisk JWR, Kemper HCG, Schuit AJ, van Mechelen W, Seidell JC. Development of overweight from adolescence into adulthood: the Amsterdam Growth and Health Longitudinal Study (AGHLS) (in press).
13. Nooyens ACJ, Visscher TLS, Verschuren WMM, Schuit AJ, van Mechelen W, Seidell JC. Age, period and cohort effects on body weight and body mass index in Dutch adults: the Doetinchem Cohort Study (in press).
14. Deurenberg P, Weststrate JA, Seidell JC. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Br J Nutr 1991;65:105–14.
15. Doak CM, Visscher TL, Renders CM, Seidell JC. The prevention of overweight and obesity in children and adolescents: a review of interventions and programmes. Obes Rev 2006;7:111–36.
16. Cole TJ, Roede MJ, Centiles of body mass index for Dutch children aged 0–20 years in 1980—a baseline to assess recent trends in obesity. Ann Hum Biol 1999; 26:303–8.

Long-term longitudinal studies and implications for the development of an international growth reference for children and adolescents

John H. Himes

Abstract

This report reviews 21 long-term, longitudinal studies of physical growth as background for the International Growth Reference for Children and Adolescents (IGRCA) initiative. Longitudinal studies form a large share of the evidence base for much of the knowledge on normal growth of children, and the collective experience from their long history is instructive relative to future studies that may result from the IGRCA. Many of the studies were initiated in the 1920s and 1930s when some current techniques, such as the use of doubly labeled water for the assessment of energy expenditure or dual-energy x-ray absorptiometry (DEXA) for the study of body composition, were not available. Nevertheless, many well-established protocols for anthropometry and for assessment of somatic maturation are as important today as they were in the past. With some important exceptions, few of the studies collected detailed information on dietary intake or child health and illness. Genetic or familial factors were limited as well. Many lessons can be drawn from the past experience with prominent longitudinal growth studies. Nevertheless, the exact design, sampling, and measurement protocols chosen for future growth studies emanating from the IGRCA effort must be carefully linked to specific research questions and the explicit purposes for which the resultant data will be used.

Key words: Growth, longitudinal study, maturation, reference data

The author is affiliated with the Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, USA.

Please direct queries to the author: John H. Himes, Division of Epidemiology and Community Health, University of Minnesota, 1300 S 2nd Street, Suite 300, Minneapolis, MN 55454-1015, USA; e-mail: himes@epi.umn.edu.

Introduction and rationale

Background

In 2003, a meeting in Rome brought together representatives from the Department of Nutrition for Health and Development at the World Health Organization (WHO), the United Nations University Food and Nutrition Program (UNU-FNP), and the Food and Agriculture Organization (FAO) to consider the feasibility and appropriateness of developing a single international growth reference or standard for school-aged preadolescents and adolescents. In particular, the attendees outlined a process for evaluating the potential content and appropriateness of an internationally applicable growth standard, presenting children's optimum growth, rather than a growth reference describing the current growth status of a particular population, some of whom might not be growing optimally. This meeting launched the International Growth Reference for Children and Adolescents (IGRCA) effort. Included in the plan of action emanating from the Rome meeting was commissioning a series of articles to review the available science critical for evaluating the feasibility, appropriateness, and potential content of an international growth reference for school-aged and adolescent children.

An obvious source of guidance concerning potential design and content of research leading to international growth references is long-term longitudinal studies of physical growth and maturation of children. Long-term longitudinal studies are important to consider, for several reasons. First, new research yielding an international growth reference may follow children longitudinally because of the importance of using growth increments or growth rates to evaluate children, and the requirement for measurements on multiple occasions to derive such estimates of growth velocity. Also, if the reference data are to extend through adolescence, an important indicator of maturational timing of the adolescent spurt, i.e., peak height velocity (PHV),

requires measurements of the same individuals on multiple occasions over a period of years. Following the same children over time may be more efficient in some situations than using cross-sectional designs, because fewer children are required to be recruited.

Long-term longitudinal studies of physical growth and maturation are particularly relevant to consider for the IGRCA, because they have been the evidence base for much of the scientific knowledge on normal growth in children, provide successful examples of longitudinal study designs, have identified important measurements to collect, have well-established measurement protocols, provide data and experience concerning meaningful measurement intervals, and have often been the basis for previous growth and maturation reference data.

Focus and scope

The focus and scope of this review serve to identify those long-term longitudinal studies of physical growth and maturation to be considered, as well as to identify aspects of the studies that are most relevant to the IGRCA. The studies reviewed are observational in nature and represent normal or typical development of individuals in the populations studied. Ideally, the studies include at least the school-age years from 5 or 6 years of age through adolescence for the same individuals; these are usually purely longitudinal studies that enrolled all children at a similar age and followed them regularly for long periods of time. In practice, there have been few such studies, and investigators often have used mixed-longitudinal designs, enrolling children at staggered ages and following them for shorter periods, while still covering a wide total age range.

Shorter-term longitudinal studies (under 4 years), those focusing on sick children or children with a specific condition (e.g., achondroplasia, cerebral palsy), and studies linked with experimental interventions or drug treatments are not considered. Studies limited only to the preschool years or with gaps between examination visits (more than 2 years) are not included. Finally, studies collecting data on only stature and weight, with less than 100 participants, or spanning less than six whole-year age groups have not been included. For example, the Berkeley Growth Study was an influential study, but it included only 61 participants [1]. In the Melbourne Longitudinal Study [2], there were insufficient details published about specific measurements for the study to be useful, even though it met the other inclusion and exclusion criteria.

An effort has been made to include those aspects and variables of the longitudinal studies that are most relevant to the goals of the IGRCA and to the development of widely applicable growth standards. For example, specialized data collections or measurements collected routinely that are unlikely to be considered for the IGRCA (e.g., dermatoglyphics, craniofacial growth, visual acuity) are not reviewed. Many longitudinal studies have included large batteries of psychological tests in order to study mental and psychosocial development. Although these are important aspects of child development, the measurements are not discussed here, because they are unlikely to be included in studies emanating from the IGRCA.

General approach

Long-term longitudinal studies of physical growth and maturation meeting the above criteria were identified from the scientific literature. An effort was made to be comprehensive, although some studies may have been overlooked. The goal was to apply the knowledge gained from longitudinal studies to the IGRCA effort, rather than to make an exhaustive list of longitudinal studies per se. Consequently, unless the omitted studies have truly unusual characteristics, any omissions should not appreciably affect the conclusions drawn. Descriptions of earlier American longitudinal studies have been published [3], as have summaries of those studies (mostly conducted on preschool children) coordinated through the International Children's Centre [4]. Tanner [5] has provided useful historical contexts to the chief longitudinal studies in America and Europe.

To simplify current reference to the longitudinal studies, selected aspects have been summarized, and each study has been assigned a number and an abbreviated name (**table 1**). The sections that follow are organized according to pertinent aspects of the studies with implications for the IGRCA program.

Study design and participants

Study dates and designs

Many of the most prominent American longitudinal studies, such as the Iowa, Third Harvard, Denver, Guidance, Fels, and Brush studies, started between 1920 and 1932 (**table 1**). The dates on which the longitudinal studies were initiated are only indirectly relevant to the present discussion, in that technology and measurement protocols may have changed over time; consequently, caution may be required in using the same methods now, because they are no longer appropriate. For example, in the Iowa study, subcutaneous fat thicknesses were measured starting in 1929 with a Franzen spring caliper [30], well before caliper jaw surface areas and tension were standardized in the Harpenden and Lange calipers that were developed in the 1950s and early 1960s [31, 32].

The relatively early dates of many of the longitudinal studies also explain why some factors now known

TABLE 1. Major longitudinal studies of growth and maturation of children

Study no.	Study name[a]	Dates	Design	Age (yr)	No. of children	Area and country	Reference
1	Iowa	1920–1934	Mixed	0–18	2,484	Iowa City, USA	Meredith, 1935 [6]; Boynton, 1936 [7]
2	3rd Harvard	1922–1934	Pure	6–17	1,553	Boston, USA	Dearborn et al., 1941 [8]
3	Denver	1927–1967	Pure	0–21	334	Denver, USA	McCammon, 1970 [9]
4	Guidance	1928–1947	Pure	0–18	136	Berkeley, USA	Tuddenham and Snyder, 1954 [10]
5	Fels	1929–present	Pure	0–21	1,036	Ohio, USA	Roche, 1992 [11]
6	4th Harvard	1930–1956	Pure	0–18	134	Boston, USA	Stuart,1939 [12]
7	Brush	1931–1942	Mixed	0–17	999	Cleveland, USA	Simmons, 1944 [13]
8	California Boys	1932–1939	Pure	11–17	233	Berkeley, USA	Stolz and Stolz, 1951 [14]
9	Harpenden	1948–1971	Mixed	3–18	420	London, UK	Tanner, 1962 [15]
10	Paris	1953–1975	Pure	0–21	542	Paris, France	Sempe et al., 1979 [16]
11	Philadelphia	1948–1968	Mixed	6–17	1,930	Philadelphia, USA	Krogman, 1970 [17]
12	Zurich	1954–1976	Pure	0–20	413	Zurich, Switzerland	Prader et al., 1989 [18]
13	West Bengal	1952–1966	Mixed	0–21	562	West Bengal, India	Das et al. 1986 [19, 20]
14	Stockholm	1955–1978	Pure	0–17	212	Stockholm, Sweden	Karlberg and Taranger, 1976 [21]
15	Wroclaw	1961–1972	Pure	8–18	470	Wroclaw, Poland	Bielicki and Waliszko, 1975 [22]; Waliszko and Jedljnska, 1976 [23]
16	Saskatchewan	1964–1973	Pure/mixed[b]	7–17	305	Saskatchewan, Canada	Mirwald, 1978 [24]
17	Leuven	1968–1974	Pure	12–18	588	Leuven, Belgium	Beunen et al., 1988 [25]
18	Nymegen	1971–1976	Mixed	4–14	467	Nymegen, Holland	Prahl-Andersen et al., 1979 [26]
19	Leeds	1972–1985	Mixed	9–18	396	Leeds, UK	Buckler, 1990 [27]
20	Mexico	1977–1980	Mixed	10–15	510	Mexico City, Mexico	Faulhaber, 1989 [28]
21	Western Australia	1981–1986	Mixed	8–17	438	Perth, Australia	Blanksby, 1994 [29]

a. Formal names have been shortened for ease of reference.
b. Pure longitudinal for boys, mixed longitudinal for girls.

to play major roles in child growth, such as physical activity and diet, were infrequently or poorly measured, and why measurements made with newer technologies such as DEXA and doubly labeled water are absent. Also, data that were collected at earlier dates in the longitudinal studies represent populations absent any secular changes that may have occurred since that time. Accordingly, direct comparisons of current data values with those from older samples should be done with caution. For example, in the Fels Longitudinal Study, the mean body-mass index (BMI) of adolescent girls was 1.19 kg/m^2 higher than that of girls born in the period from 1929 to 1946, and the girls born between 1965 and 1983 had a correspondingly greater total increase in BMI during adolescence [33].

By definition, longitudinal studies collect multiple observations on children, and in the present case, they do so over long periods of time. As seen in **table 1**, the major longitudinal studies have used both purely longitudinal and mixed-longitudinal designs. Usually, studies of very large numbers of children, such as the Iowa and Philadelphia studies, have mixed-longitudinal designs. Specific implications of various study designs for the IGRCA are addressed by Cole in a separate contribution to this issue [34].

Visit schedules and intervals between measurements

The timing of scheduled examinations of participants and the intervals between them have several implications for uses and interpretation of the resultant data. During periods of rapid growth, more frequent examination visits allow a more complete description of the patterns of growth, although costs and logistical difficulties increase accordingly. The frequencies of examination visits for selected one-year periods are summarized in **table 2**. Within each of the one-year periods, the visits were usually scheduled at equal intervals; for example in the Stockholm study, the four visits from birth to 0.99 years of age were scheduled at 1, 3, 6, and 9 months. The frequencies of examination visits are higher during infancy and adolescence, which are periods of rapid growth.

The optimum timing and frequency of examination visits in a longitudinal study should be dictated primarily by the intended applications of the resultant data, and secondarily by the available labor and financial resources. If a chief purpose of an IGRCA longitudinal study is to produce growth reference data relative to age during middle childhood and adolescence, the examination visits during adolescence must be frequent enough to characterize the abrupt changes in growth velocity during the adolescent spurt and to identify accurately the age at peak height velocity in individuals. This level of description requires examinations at intervals of no longer than 6 months, and preferably at intervals of 3 to 4 months during the full range of years when almost all children will go through adoles-

TABLE 2. Frequency of examination visits for measurement of stature and weight within selected age periods of one year for longitudinal studies

Study no.	Study name	0–0.99 yr	1.0–1.99 yr	5.0–5.99 yr	8.0–8.99 yr	12.0–12.99 yr	14.0–14.99 yr	17.0–17.99 yr
1	Iowa	4	4	2	1	1	1	1
2	3rd Harvard	—	—	1	1	1	1	1
3	Denver	7	2	2	2	2	2	2
4	Guidance	4	4	1	2	2	2	2
5	Fels	4	2	2	2	2	2	2
6	4th Harvard	5	2	2	2	1	1	1
7	Brush	3	2	1	1	1	1	1
8	California Boys	—	—	—	—	2	2	2
9	Harpenden	—	—	2	2	4	4	4
10	Paris	5	4	2	2	2	2	2
11	Philadelphia	—	—	—	1	1	1	1
12	Zurich	5	2	1	2	2	2	1
13	West Bengal	1	2	1	1	2	2	1
14	Stockholm	4	2	1	1	4	4	4
15	Wroclaw	—	—	—	1	1	1	1
16	Saskatchewan	—	—	—	1	1	1	1
17	Leuven	—	—	—	—	1	2	2
18	Nymegen	—	—	1	1	1	1	—
19	Leeds	—	—	—	1	3	3	3
20	Mexico	—	—	—	—	2	2	—
21	Western Australia	—	—	—	2	2	2	1

cence. During the years preceding adolescence, annual examinations are probably sufficient to adequately capture the patterns of growth.

It is important to note that the optimum intervals between examination visits for a longitudinal study that will use data in the aggregate for a growth reference are not necessarily the same as the optimum intervals for detecting growth in individual children. In the former application, the intended task is to estimate means and quantiles of attained growth and growth velocity, and to describe the characteristic patterns of changes in growth velocity with reasonable precision and accuracy. The precision and accuracy of means and quantiles of size and velocity are primarily dependent on the number of children and the sampling frame at a given age, whereas the accurate description of changes in growth velocity depends on having sufficiently frequent examination visits to capture the features of the growth curve. In contrast, the optimum intervals between serial measurements of individual children to detect meaningful growth depend on the reliability of the measurements, the normal variation in child size on the target ages, and the expected rates of growth [35]. Because attained size and rates of growth vary systematically according to age, so do the minimal time intervals between examinations that are necessary to detect growth.

Some of the longitudinal studies provide acceptable tolerances for the target ages for examination visits. For example, visit tolerances from the Fels longitudinal study are presented in **table 3** [11]. From the 6-month visit until the visit at 7.5 years, the tolerance is about 1% of chronological age; tolerance then decreases as the child gets older. The small tolerances for scheduled examinations provide data points very close to prescribed ages and make age-specific reporting convenient. Furthermore, small tolerances minimize age-related covariance and the resulting inflated standard deviations of growth variables in age-specific samples. Nevertheless, some examinations invariably will occur outside of the prescribed tolerances. In the past, to ensure the precision of age-group definitions, these off-schedule data may not have been included in reports. More recently, investigators have used the serial data for individuals and mathematical functions to interpolate measurements at exact ages when needed, for example, for calculation of increments [36, 37], or have used statistical approaches for group analyses that can accommodate differences in age at examination [11].

TABLE 3. Tolerances (± no. of days) for target ages at examination at the Fels Longitudinal Study

Target age	Tolerance (days)
≤ 6 mo	2
9 mo	3
1.0 yr	4
1.5 yr	5
2.0 yr	7
2.5 yr	9
3.0 yr	11
3.5 yr	12
4.0 yr	14
4.5 yr	16
5.0 yr	17
5.5 yr	18
6.0 yr	22
6.5 yr	23
7.0 yr	25
≥ 7.5 yr	30

Source: Roche [11].

Participants

It is difficult to ensure that participants in long-term longitudinal studies are statistically representative of national populations because of the relatively small samples and the requirement to retain participants who live in convenient locations for multiple clinic visits over long periods. Accordingly, individuals with long series of measurements are a self-selected subsample of all participants.

The participants in most of the longitudinal studies are from families of middle or upper-middle socioeconomic status, defined according to local standards at the time. Nevertheless, in several studies, such as those from Leuven, Saskatchewan, Wroclaw, and Stockholm, care was taken to sample multiple socioeconomic levels or even to explicitly stratify samples according to socioeconomic status.

Almost all of the longitudinal studies have primarily participants who are white children of European heritage. The exceptions to this generalization are the Philadelphia study, with its planned inclusion of African-American children, and the West Bengal and Mexico samples, which included local children who usually would be considered Indian or Latino, respectively. A few African-American participants were included in the Third Harvard and the Fels longitudinal studies, but their data usually have not been reported.

The details concerning recruitment of study participants at or near the time of birth into the longitudinal studies are sketchy, although prenatal clinics were the source for the Fourth Harvard and the Stockholm studies, and sampling from birth records is indicated for the Iowa and Guidance studies. Several of the studies, including the Third Harvard, Guidance, Mexico, Wroclaw, Leeds, Philadelphia, and California Boys studies, recruited participants from local schools. The Harpenden study is unique among the longitudinal studies in sampling residents of a children's home (near London).

For the IGRCA, the key issue in identifying participants for a longitudinal study leading to development of a growth reference will be the specific intent of the final product. If the resulting growth reference is prescriptive and provides a standard of optimum growth, much attention must be given to definition and selection of the children, their families, and even communities to meet the conceptual desiderata. If the resulting growth reference is to provide current descriptions of the growth of children in naturally occurring populations, participants will need to be identified through sampling frames that are statistically representative. It will be logistically crucial to enroll children (and their families) in ways that will ensure their continued participation and retention in the study cohorts.

Chief measures and data collected

Anthropometry

The anthropometric dimensions measured most commonly in the longitudinal studies (in at least three studies) are presented in **table 4**. Weight and stature were the only measurements included in every study at all ages. Recumbent length was measured in all studies that included examination visits for infants and children less than 3 years of age. Several studies (the Fels, Fourth Harvard, Brush, and Zurich studies) included recumbent length measurements at every visit throughout childhood and adolescence, or at least through middle childhood (the Stockholm study). Certainly, weight and stature are the anthropometric dimensions

TABLE 4. Anthropometric dimensions measured most commonly in the longitudinal studies (identified by study number according to table 1)

Anthropometric dimension	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Weight	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Heights																					
Stature	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Sitting	■	■	■	—	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Subischial[a]	—	■	■	—	■	■	■	—	■	■	■	■	—	■	—	■	—	—	■	■	■
Lengths																					
Recumbent	■	—	■	—	■	■	■	—	■	■	—	■	■	■	—	—	—	—	—	—	—
Crown-rump	—	—	■	■	—	■	■	■	■	—	—	■	—	■	—	—	—	—	—	—	—
Arm[a]	—	■	■	—	■	—	■	—	—	—	■	■	■	—	■	—	—	—	—	■	■
Breadths																					
Biacromial	■	—	■	■	■	—	■	■	■	■	■	—	—	■	■	■	■	■	■	■	■
Biiliac	—	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	—	■	■	■	—
Bitrochanteric	■	—	—	—	—	—	■	■	—	—	■	—	—	—	■	—	—	—	—	—	■
Knee	■	—	■	—	■	—	—	—	■	■	—	■	—	■	■	■	■	■	■	■	■
Elbow	■	—	■	—	■	■	—	—	■	■	—	■	—	■	■	■	■	■	■	■	■
Chest	■	■	■	—	—	■	■	■	—	—	■	—	■	■	■	■	■	—	—	—	—
Head	■	■	■	—	—	■	■	—	—	—	—	■	■	—	■	—	—	—	—	■	—
Circumferences																					
Head	—	■	■	—	■	■	—	—	—	■	—	■	—	■	—	—	—	■	—	■	—
Chest	■	—	■	—	■	■	■	■	—	■	—	■	—	■	■	■	■	—	—	■	■
Thigh	■	—	—	—	■	—	—	■	■	—	—	—	—	—	■	■	■	—	—	—	■
Calf	■	—	■	■	■	—	—	■	■	■	—	■	—	■	—	■	■	—	■	—	■
Arm	■	—	■	—	■	—	—	■	■	■	—	■	—	■	■	■	■	■	■	■	■
Abdominal	—	—	■	—	■	■	—	■	—	—	—	—	—	—	—	—	—	—	—	—	■
Hip	—	—	—	—	■	—	—	—	—	—	—	—	—	—	■	■	—	—	—	—	—
Skinfolds																					
Triceps	■	—	■	—	■	—	—	—	■	■	—	■	—	■	■	■	■	■	■	■	■
Biceps	■	—	■	—	■	—	—	■	■	■	—	■	—	■	—	—	—	■	■	—	—
Subscapular	■	—	■	—	■	—	—	—	■	■	—	■	—	■	■	■	■	■	■	■	■
Suprailiac	■	—	■	—	■	—	—	■	■	■	—	■	—	■	—	■	■	■	■	■	■
Calf	—	—	■	—	■	—	—	—	—	—	—	—	—	—	—	■	■	—	—	■	■
Abdominal	—	—	■	—	—	—	—	■	—	—	—	—	—	—	■	■	—	—	—	—	■

a. The dimension may be measured directly or derived from other measurements.

used most widely in clinical and public health settings for general evaluation of healthy development and as indicators of undernutrition, subsequent health risk, and overweight and obesity [38].

Sitting height was measured in all studies except for the Guidance study, where crown–rump length, which provides very similar information, was measured. The only segment length commonly measured was arm length.

Body breadths or widths have been used as measures of frame size and as anthropometric predictors of body composition [39]. The body breadths most commonly included in the longitudinal studies were biacromial (shoulder) and biiliac (hip) breadths; measurements of bony breadths at the knee and elbow were almost as common. Although ankle and wrist breadths were measured rarely in the longitudinal studies (only in the Third Harvard and Denver studies), they have been shown, at least in adults, to be good predictors of lean mass while being virtually uncorrelated with body fat; consequently, they are excellent candidates to aid in discriminating between fat and lean [40].

The most commonly measured circumferences in the longitudinal studies were those of the arm, calf, and chest. Arm and calf circumferences have become important anthropometric indicators of undernutrition [38], whereas chest circumference was commonly used as a measure of frame size in the past [39]. Head circumference is routinely used for clinical evaluation of children but was not always included in the longitudinal studies, even during the first few years of growth when head size is of most interest to clinicians.

Few of the longitudinal studies routinely measured waist or abdominal circumference and hip circumference. Almost all of the longitudinal studies, however, were initiated (and many completed) before the first general recognition of the importance of waist and hip circumferences as measures of visceral and subcutaneous fat distribution and as indicators of health risk [41]. Certainly, any new longitudinal studies that consider cardiovascular or diabetic risk factors should include waist and hip circumferences to document the development of these dimensions that have become so important as indicators of subsequent health risk.

Skinfold thicknesses measure the double thickness of compressed skin and subcutaneous fat at particular body sites. Subcutaneous fat has received attention because of its accessibility with noninvasive methods and because it is highly correlated with total body fat [42]. Even though skinfold protocols and calipers were not fully standardized until after 1950, skinfolds were included in many of the longitudinal studies. The most common sites for skinfold measurements were the triceps, subscapular, and suprailiac sites, which have been recommended repeatedly because of reliability, validity, and risk prediction [43]. In the Fourth Harvard, Denver, and Fels longitudinal studies, subcutaneous fat thicknesses were measured without compression directly from soft-tissue radiographs taken at several trunk and extremity sites, including deltoid, forearm, thigh, calf, 10th rib, and hip [44]. At Fels and Saskatchewan, the traditional measures of fatness and body composition have been augmented by DEXA and bioelectric impedance in recent years.

For some of the longitudinal studies, growth variables routinely derived from direct measurements were specified. These derived variables are generally of two sorts, projected segment lengths and ratios or indexes. The anthropometric methods passed on from the nineteenth century measured body segment lengths as *projected* differences between theoretical horizontal planes at the levels of bony landmarks that were measured as heights from the floor. For example, projected arm length was derived as the difference between the height of acromion (the most lateral point of the acromial process of the scapula) and the height of dactylion (the most distal point of the third finger with the arm hanging naturally at the side). Several of the longitudinal studies, such as the Third Harvard, Fels, Brush, Philadelphia, and Mexico studies, included some of these traditional landmark heights and used them to derive segment lengths. Subischial length (or height), the difference between stature and sitting height, is one of the few vestiges of this traditional approach of projected lengths that is commonly included in anthropometric protocols nowadays [46].

Several longitudinal studies specifically mentioned derived measures that were routinely calculated as ratios or indexes: e.g., stature:sitting height, biacromial breadth:biiliac breadth, sum of skinfolds, and body-mass index (weight in kilograms divided by the square of stature in meters). Obviously, these derived variables may have been calculated from the same direct measurements and used routinely in other studies as well, but not mentioned in the study descriptions.

Measures of maturation

Measures of maturational status collected in the longitudinal studies are presented in **table 5**. Skeletal maturation was assessed in almost all of the longitudinal studies. Skeletal maturation is especially attractive because it provides quantitative estimates of maturational progress from infancy through much of adolescence [47]. Measurement of maturational status based on a radiograph of the hand and wrist was the most commonly used method in the longitudinal studies, reflecting the early work of Baldwin et al. [48] and Todd [49], which was later extended by Greulich and Pyle [50], Tanner et al. [51], and Roche et al. [52]. Skeletal maturation assessments based on radiographs at other joint sites were included mainly in response to the early scoring systems proposed by Todd that were subsequently developed fully only for the knee [47, 53] and the foot-ankle [54].

The ages at eruption or emergence of deciduous and

TABLE 5. Measures of maturational status in the longitudinal studies (identified by study number according to table 1)

Maturation measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Skeletal																					
Hand-wrist	■	■	■	■	■	■	■	■	■	■	—	—	■	■	■	■	—	■	■	—	—
Knee	—	—	—	■	■	■	■	■	—	■	—	—	—	—	—	—	—	—	—	—	—
Elbow	—	—	—	—	—	■	■	—	—	■	—	—	—	—	—	—	—	—	—	—	—
Foot-ankle	—	—	—	—	—	■	■	—	—	■	—	—	—	—	—	—	—	—	—	—	—
Shoulder	—	—	—	—	—	■	■	—	—	■	—	—	—	—	—	—	—	—	—	—	—
Hip	—	—	—	—	—	—	■	—	—	■	—	—	—	—	—	—	—	—	—	—	—
Dental																					
Deciduous eruption	—	—	—	—	■	■	—	—	■	■	—	—	—	■	—	—	—	■	—	—	—
Permanent eruption	—	■	—	—	■	■	—	■	■	■	—	—	—	■	—	—	—	■	—	—	—
Sexual—male																					
Genital	—	—	—	—	■	—	—	■	—	■	—	■	—	■	—	—	—	■	■	—	■
Pubic hair	—	—	—	—	■	—	—	■	—	■	—	■	—	■	—	—	—	■	■	—	■
Axillary hair	—	—	—	—	—	—	—	■	—	—	—	—	—	■	—	—	—	—	—	—	—
Testicular volume	—	—	—	—	—	—	—	—	—	■	—	■	—	■	—	—	—	■	■	—	—
Voice change	—	—	—	—	—	—	—	—	—	—	—	—	—	■	—	—	—	—	—	—	—
Sexual—female																					
Breast	—	—	—	—	■	—	—	—	—	■	—	■	—	■	—	—	—	■	■	—	■
Pubic hair	—	—	—	—	■	—	—	—	—	■	—	■	—	■	—	—	—	■	■	—	■
Axillary hair	—	—	—	—	—	—	—	—	—	—	—	—	—	■	—	—	—	—	—	—	—
Menarche	—	■	—	—	■	—	—	—	—	■	—	■	—	■	■	—	—	■	■	—	—

permanent teeth or the number of teeth erupted at a given age can be used as measures of dental maturation [55]. A few of the longitudinal studies collected this information on dental development, and at Fels additional measures of individual tooth maturation were assessed from panoramic radiographs. Eruption of the primary dentition is primarily controlled genetically and is extremely resistant to environmental factors [56]. Eruption of the permanent dentition, or the number of permanent teeth at a given age, is a useful measure of somatic maturation for groups but is insufficiently sensitive to be useful for individual-level applications [57].

The development of secondary sexual characteristics was assessed in many of the longitudinal studies as an indicator of sexual maturation. Descriptive stages of the qualitative changes in type and distribution of body hair and of the development of genitalia in males and breasts in girls had been proposed before most of the early longitudinal studies were initiated [58], but the stages were not standardized and routinely recorded until after the work of Nicolson and Hanley [59] and Reynolds and Wines [60, 61]. After these stages of sexual maturation were made more available by Tanner [15], most of the longitudinal and other growth studies adopted his stages, and in some cases even have referred to them as Tanner stages.

Testicular volume, assessed by palpation and comparisons with models of known volume, was used as a measure of sexual maturation in some of the longitudinal studies that were begun after 1950. Voice change in boys was routinely recorded only in the Stockholm longitudinal study, but the reliability and validity of this indicator were quite good [21]. Most of the longitudinal studies recorded menarcheal status as an indicator of sexual maturation in girls.

An important indicator of maturational timing derived in many of the longitudinal studies was age at PHV or peak stature velocity, the estimated age of maximum velocity of growth in stature during the adolescent spurt. As a cautionary note, the timing and intervals between measurements during the adolescent spurt and the method used to estimate the peak age can result in surprisingly large differences [11].

For growth studies that are developed as part of the IGRCA, measures of somatic maturation should be carefully considered, especially if the adolescent years are included. During adolescence, significant variation in growth variables is associated with maturational status, even within strictly defined chronological age groups [62]. Moreover, the times of maturational thresholds or landmarks, such as peak height velocity or menarche, have become important descriptors of the tempo of development [38].

The exact maturational measurements and indicators used for IGRCA studies must depend on the specific purposes of their use, technical and personnel requirements for their implementation, and the acceptability of the measurements [38]. For example, although the ionizing radiation associated with a hand-wrist radiograph is minimal, it may not be acceptable to some national or local policies concerning radiation safety. Direct observation of secondary sex characteristics may not be acceptable to some communities or individuals, even if conducted by observers of the same sex.

Hematologic and biochemical measurements

Few of the longitudinal studies routinely drew blood samples or obtained urine. Exceptions to this pattern were the Denver and Nymegen longitudinal studies. At Denver, blood was obtained at the regularly scheduled examination visits and was assayed for what we would now call a complete blood count (CBC) and also for sedimentation rate, total cholesterol, alpha-lipoprotein cholesterol, beta-lipoprotein cholesterol, total protein, fibrinogen, albumin, and alpha-, beta-, and gamma-globulins. At Nymegen, regular hematologic measurements included hemoglobin, hematocrit, erythrocyte count, serum iron, iron-binding capacity, transferrin saturation, immunoglobulin concentration, and gonadotropin concentration.

At Fels, serum alkaline phosphatase, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and urinary creatine and creatinine were measured for a number of years. Hemoglobin and erythrocyte count were measured routinely in the Fourth Harvard study, and urinary gonadotropins were measured on four annual occasions in the Leeds study.

Hematologic and biochemical measures may be considered for IGRCA studies for many purposes, including selection criteria for participation in the studies, description of the general health and nutritional status of the participants, indicators of sexual maturation, or as outcome variables of interest relative to specific research questions. The previous longitudinal studies offer little guidance on specific measurements that should be taken, but experience with the Denver and Nymegen studies indicates that routine sampling for hematologic and biochemical measurements can be successfully incorporated into long-term longitudinal efforts.

Measures of feeding or diet

As a part of the Fourth Harvard longitudinal study, Bertha Burke developed the dietary history method [63] of recording the kind, amount, and frequency of foods consumed during a period of 1, 3, or 6 months, covering the interval since the previous interview. Burke trained Virginia Beal, who directed the nutrition component in the Denver study. The Denver study used the Burke dietary history, as well as four 24-hour recalls on different days close to each scheduled examination visit [64]. Analyses of the dietary data from these two longitudinal studies were very influential, and they were highly unusual at the time because of their longitudinal nature. At Fels, maternal diet (e.g., Sontag et al. [65]) and breastfeeding and diet in infancy [66] were studied for a period of years in the longitudinal study participants. Nevertheless, few detailed dietary data were collected in the other longitudinal studies as part of their routine long-term protocols.

The dietary history and 24-hour recall methods are still often used in forms fundamentally unchanged from the way they were applied in the longitudinal studies, although there have been major advances in the nutrient and food databases used with them. Nevertheless, whether and how dietary intake per se should be measured for IGCRA-related studies, again, must depend on the specific research questions to be answered by the studies.

Physical activity, fitness, and exercise

Physical activity per se or proxy measures for energy expenditure by participants were not of major interest in the longitudinal studies; indeed, even in the general scientific literature, these concerns only became more prominent in the last quarter of the past century. The Saskatchewan longitudinal study originated in a department of physical education and kinesiology at a comparatively late date (1964), so it is not surprising that questionnaires assessing physical activity and sports participation were routinely applied. Nevertheless, similar data were not collected in the other longitudinal studies, even those initiated more recently and housed in similar academic departments, such as the Leuven and West Australia studies.

Several of the longitudinal studies (i.e., the Iowa, Guidance, California Boys, Leuven, Nymegen, and West Australia studies) included one or more measures of static strength, such as grip strength, arm pull, or leg pull. The Saskatchewan, Leuven, and West Australia studies also included measures of explosive and functional strength, flexibility, speed of limb movement, and running speed. The Saskatchewan and West Australia studies also measured pulmonary function and working capacities. Clearly, the longitudinal studies focused more on the development of functional outcomes of the participants, rather than viewing physical activity as an exposure, as one would in more epidemiologically oriented studies focusing on nutritional status or other health outcomes.

Medical and health status

Routine health histories and physical examinations are listed for relatively few of the longitudinal studies, namely, the Denver, Fourth Harvard, California Boys, Nymegen, and Saskatchewan studies. Nevertheless,

because several of the longitudinal studies (e.g., the Paris, Zurich, and Stockholm studies) were housed in and operated by departments of pediatrics, it seems likely that data from health histories and physical examinations were routinely collected, even if the published accounts of the longitudinal studies do not list them. Actually, very few health data from the longitudinal studies have been published. Exceptions to this include the data on health histories and electrocardiography from the Denver study [9] and the landmark studies of longitudinal illness experience from the Fourth Harvard study [67]. Alex Roche completed a massive review (1,400 pages) of the development of blood pressure in children, including analyses of the blood pressure data from the Fels, Fourth Harvard, Denver, and Berkeley longitudinal studies; unfortunately, these analyses are only available in an unpublished technical report to the US National Institutes of Health.*

In general, it appears that the main purpose of most of the health examinations in the longitudinal studies was to document and ensure the health of the participants, rather than to collect health data to be used as outcome variables to answer specific research questions.

Family measurements and genetic aspects

Few of the longitudinal studies routinely collected data from parents or families of participants, other than data necessary for demographic characterization of the samples or data on prenatal factors during the pregnancy resulting in the participant's birth. Parental stature and weight were measured in the Denver, Fels, Fourth Harvard, and Philadelphia studies. In the Fourth Harvard, Denver, and Fels studies, some siblings were included in the samples, although for some analyses only one child per family was included.

At Fels, recruitment focused on families, not just on individuals, so that up to three generations of the same families have been participants in the longitudinal study [11]. The parent–child relationships at Fels have allowed many familial analyses [68], the use of parental data in a method to predict adult stature [69], and a method to adjust childhood stature for that of parents [70]. The pedigrees of Fels families have also been used in quantitative genetic analyses [71].

Concluding comments concerning implications for the IGRCA

The longitudinal studies considered span a wide range of time and place. It is not at all clear that any of the longitudinal studies were designed primarily, or even secondarily, for the purpose of developing a growth reference or standard, even though data from some of the studies subsequently have been used for that purpose [16, 36, 37, 72].

Most of the longitudinal studies focused on describing normal physical growth and maturation and were less concerned about child health or nutrition per se, which are major concerns for the IGCRA. Exceptions to this generalization are the Fourth Harvard and Denver studies. The longitudinal studies did, however, reflect the academic traditions and training of their chief investigators. Many of the earliest longitudinal studies, such as the Iowa, Third Harvard, Guidance, Brush, and California Boys studies, originated out of comprehensive views of child development in psychology, education, and sociology that included physical development. Some of the studies starting later, including the Philadelphia, West Bengal, Wroclaw, Saskatchewan, Leuven, Mexico, and Western Australia studies, arose out of traditions in anthropology and physical education.

There are some additional issues related to longitudinal studies that are seldom mentioned explicitly in descriptions of the studies but that are still very important. First is the need for carefully standardized measurement protocols and well-trained data collectors. Because serial data for individuals will be used, perhaps including calculated increments of change, both random measurement errors and measurement bias need to be minimized as much as possible [73]. Maintaining a small number of anthropometrists through time and tracking interobserver variation in measurements favor high-quality measurements. Standardized data-cleaning and management ensure that unusual values, late visits, and missing data are handled in a similar manner over time.

Because longitudinal studies involve the same participants and families over an extended period, special effort is required by study staff to foster good relationships with the participants and families to ensure timely visits and retention of participants. Special attention to participant burden and family circumstances is a worthwhile endeavor and is paid for many times over in terms of participant good will, cooperation, and community acceptance of the study.

Growth studies that may result from the IGRCA will provide an opportunity to design research specifically to yield data suitable for the development of a growth reference or standard for childhood and adolescence. By carefully defining research questions and specific purposes for the data and then linking the design, sampling, and specification of variables closely to these purposes, the accumulated knowledge and experience of the previous longitudinal studies can provide wise lessons for the future.

* Roche AF, Eichorn D, McCammon RW, Reed RB, Valadian I, Himes JH, Kent RLJ, Siervogel RM. The natural history of blood pressure. Bethesda, MD, USA: National Heart, Lung and Blood Institute, National Institutes of Health, 1980.

Acknowledgments

Matthew Varner, an MPH student at the University of Minnesota, tracked down references and study details and developed tables summarizing the longitudinal studies using dimension names from Lohman et al. [46]. When the well of potential studies to consider seemed dry, I drew from the encyclopedic knowledge of Bob Malina, who graciously filled in gaps to rehydrate the array of possibilities.

References

1. Macfarlane JW. Objectives, samples and procedures. In: Jones MC, Bayley N, Macfarlane JW, Honzik MP, eds. The course of human development: selected papers from The Longitudinal Studies, Institute of Human Development, The University of California, Berkeley. Waltham, Mass, USA: Xerox College Publishing, 1971.
2. Roche AF, Sunderland S. Melbourne University Child Growth Study. Med J Aust 1959;46:559–62.
3. Tanner JM. A guide to American growth studies. Yearb Phys Anthropol 1948;3:28–33.
4. Falkner F, Vesin P, Himes JH, eds. Growth and development of the child: 25 years of internationally coordinated activities. Courrier 30 (special issue). Paris: International Children's Centre, 1980.
5. Tanner JM. A history of the study of human growth. Cambridge, UK; New York: Cambridge University Press, 1981.
6. Meredith HV. The rhythm of physical growth: a study of eighteen anthropometric measurements on Iowa City white males ranging in age between birth and eighteen years. University of Iowa Studies in Child Welfare 1935; 2(3):1–128.
7. Boynton B. The physical growth of girls: a study of the rhythm of physical growth from anthropometric measurements on girls between birth and eighteen years. University of Iowa Studies in Child Welfare 1936;12(4):1–105.
8. Dearborn WF, Rothney JWM, Long HH, Ratcliff JM, Crissy WJE. Predicting the child's development. 2nd revised ed. Cambridge, Mass, USA: Sci-Art Publishers, 1941.
9. McCammon RW. Human growth and development. Springfield, Ill, USA: Charles C Thomas, 1970.
10. Tuddenham RD, Snyder MM. Physical growth of California boys and girls from birth to eighteen years. Publ Child Dev Univ Calif 1954;1:183–364.
11. Roche AF. Growth, maturation, and body composition: the Fels Longitudinal Study, 1929–1991. New York: Cambridge University Press, 1992.
12. Stuart HC. Studies from the Center for Research in Child Development, School of Public Health, Harvard University, I. The Center, the group under observation, sources of information and studies in progress. Monogr Soc Res Child Dev 1939;4:1–98.
13. Simmons K. The Brush Foundation study of child growth and development. II. Physical growth and development. Monogr Soc Res Child Dev 1944;9(1), serial no. 37.
14. Stolz HR, Stolz LM. Somatic development of adolescent boys; a study of the growth of boys during the second decade of life. New York: Macmillan, 1951.
15. Tanner JM. Growth at adolescence, 2nd ed. Oxford, UK: Blackwell Scientific Publications, 1962.
16. Sempe M, Pedron G, Roy-Pernot M-P. Auxologie: methode et sequences. Paris: Laboratoire Theraplix, 1979.
17. Krogman WM. Growth of head, face, trunk, and limbs in Philadelphia white and Negro children of elementary and high school age. Monogr Soc Res Child Dev 1970;35(3):1–80.
18. Prader A, Largo RH, Molinari L, Issler C. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. Helv Paediatr Acta Suppl 1989;52:1–125.
19. Das SR, Mukherjee DP, Bose L, Das A. Mixed-longitudinal growth data for 22 measures: the Sarsuna Barisha series. Vol. I: Boys. Calcutta: Anthropological Survey of India, Government of India, 1986.
20. Das SR, Mukherjee DP, Bose L, Das A. Mixed-longitudinal growth data for 22 measures: The Sarsuna-Barisha series. Vol. II: Girls. Calcutta: Anthropological Survey of India, Government of India, 1986.
21. Karlberg P, Taranger J. The somatic development of children in a Swedish urban community: a prospective longitudinal study. Acta Paediatr Scand Suppl 1976;(258):1–148.
22. Bielicki T, Waliszko A. Wroclaw Growth Study: Part I: Females. Stud Phys Anthropol 1975;2:53–83.
23. Waliszko A, Jedljnska W. Wroclaw Growth Study: Part II: Males. Stud Phys Anthropol 1976;3:27–48.
24. Mirwald RL. The Saskatchewan Growth and Development Study. In: Ostyn M, Beunen GP, Simons J, eds. Kinoanthropometry II. Baltimore, Md, USA: University Park Press, 1978:278–305.
25. Beunen G, Malina RM, Van't Hof MA, Simons J, Ostyn M, Renson R, Van Gerven D. Adolescent growth and motor performance: a longitudinal study of Belgian boys. HKP Sport Science Monograph Series. Champaign, Ill, USA: Human Kinetics Books, 1988.
26. Prahl-Andersen B, Kowalski CJ, Heydendael PH. A mixed longitudinal study of growth and development. London: Academic Press, 1979.
27. Buckler JMH. A longitudinal study of adolescent growth. London and New York: Springer-Verlag, 1990.
28. Faulhaber J. Crecimiento: somatometría de la adolesciencia. Ciudad Universitaria, Mexico: Universidad Nacional Autonoma de Mexico, 1989.
29. Blanksby BA. Athletics, growth, and development in children: the University of Western Australia study. Camberwell, Victoria, Australia, and Langhorne, Pa, USA: Harwood Academic Publishers, 1994.
30. Franzen R. Physical measures of growth and nutrition. School Health Res Monogr 2. New York: American Child Health Association, 1929.
31. Edwards DA, Hammond WH, Healy MJ, Tanner JM, Whitehouse RH. Design and accuracy of calipers for

measuring subcutaneous tissue thickness. Br J Nutr 1955;9:133–43.
32. Lange KO, Brozek J. A new model of skinfold caliper. Am J Phys Anthropol 1961;19:98–9.
33. Demerath EW, Li J, Sun SS, Chumlea WC, Remsberg KE, Czerwinski SA, Towne B, Siervogel RM. Fifty-year trends in serial body mass index during adolescence in girls: the Fels Longitudinal Study. Am J Clin Nutr 2004;80:441–6.
34. Cole TJ. The international growth standard for school-aged children and adolescents: Statistical considerations. Food Nutr Bull 2006;27(suppl):250–6.
35. Himes JH. Minimum time intervals for serial measurements of growth in recumbent length or stature of individual children. Acta Paediatr 1999;88:120–5.
36. Roche AF, Guo S, Moore WM. Weight and recumbent length from 1 to 12 mo of age: reference data for 1-mo increments. Am J Clin Nutr 1989;49:599–607.
37. Roche AF, Himes JH. Incremental growth charts. Am J Clin Nutr 1980;33:2041–52.
38. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser 1995;854:1–452.
39. Himes JH. Considering frame size in nutritional assessment. In: Himes JH, ed. Anthropometric assessment of nutritional status. New York: Wiley-Liss, 1991:141–150.
40. Himes JH, Bouchard C. Do the new Metropolitan Life Insurance weight-height tables correctly assess body frame and body fat relationships? Am J Public Health 1985;75:1076–9.
41. Kissebah AH, Vydelingum N, Murray R, Evans DJ, Hartz AJ, Kalkhoff RK, Adams PW. Relation of body fat distribution to metabolic complications of obesity. J Clin Endocrinol Metab 1982;54:254–60.
42. Himes JH. Subcutaneous fat thickness as an indicator of nutritional status. In: Greene LS, Johnston FE, eds. Social and biological predictors of nutritional status, physical growth, and neurological development. New York: Academic Press, 1980:9–32.
43. Roche AF, Abdel-Malek AK, Mukherjee D. New approaches to clinical assessment of adipose tissue. Body-composition assessments in youth and adults: Report of the Sixth Ross Conference on Medical Research. Columbus, Ohio, USA: Ross Laboratories, 1985:14–9.
44. Roche AF, Siervogel RM, Chumlea WC, Reed RB, Valadian RI, Eichorn D, McCammon RW. Serial changes in subcutaneous fat thickness of children and adults. Monogr Paediatr 1982;17:1–110.
45. Wilder HH. A laboratory manual of anthropometry. Philadelphia, Pa, USA: P. Blakiston's Son & Co., 1920.
46. Lohman T, Roche A, Martorell R, eds. Anthropometric standardization reference manual. Champaign, Ill, USA: Human Kinetics Books, 1988.
47. Roche AF, Wainer H, Thissen D. Skeletal maturity: the knee joint as a biological indicator. New York: Plenum Medical Book Co., 1975.
48. Baldwin BT, Busby LM, Garside HV. Anatomic growth of children; a study of some bones of the hand, wrist, and lower forearm by means of roentgenograms. University of Iowa Studies in Child Welfare 1928;14:1–88.
49. Todd TW. Atlas of skeletal maturation. St. Louis, Mo, USA: Mosby, 1937.
50. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. Stanford, Calif, USA: Stanford University Press, 1950.
51. Tanner JM, Whitehouse RH, Cameron N, Marshall WA, Healy MJR, Goldstein H. Assessment of skeletal maturity and prediction of adult height (TW2 Method), 2nd ed. London and New York: Academic Press, 1983.
52. Roche AF, Chumlea W, Thissen D. Assessing the skeletal maturity of the hand-wrist: Fels method. Springfield, Ill, USA: Charles C Thomas, 1988.
53. Pyle SI, Hoerr NL. A radiographic standard of reference for the growing knee. Springfield, Ill, USA: Charles C Thomas, 1969.
54. Hoerr NL. Radiographic atlas of skeletal development of the foot and ankle, a standard of reference. Springfield, Ill, USA: Charles C Thomas, 1962.
55. Hagg U, Taranger J. Dental development, dental age and tooth counts. Angle Orthod 1985;55:93–107.
56. Delgado H, Habicht JP, Yarbrough C, Lechtig A, Martorell R, Malina RM, Klein RE. Nutritional status and the timing of deciduous tooth eruption. Am J Clin Nutr 1975;28:216–24.
57. Himes JH. Why study child growth and maturation? In: Hauspie R, Cameron N, Molinari L, eds. book title. New York: Cambridge University Press, 2004:3–26.
58. Stratz CH. Der Körper des Kindes, 2. Stuttgart, Germany: Verlag Ferdinand Enke, 1904.
59. Nicolson AB, Hanley C. Indices of physiological maturity: derivation and interrelationships. Child Dev 1953;24:3–38.
60. Reynolds EL, Wines J. Individual differences in physical changes associated with adolescence in girls. AMA Am J Dis Child 1948;75:329–50.
61. Reynolds EL, Wines JV. Physical changes associated with adolescence in boys. AMA AMA Am J Dis Child 1951;82:529–47.
62. Himes JH. Applications of maturational variables in public health and epidemiology. In: Gilli G, Schell L, Benso L, eds. Human growth from conception to maturity. London: Smith-Gordon, 2002:107–12.
63. Burke BS. The dietary history as a tool in research. J Am Diet Assoc 1947;23:1041.
64. Beal VA. Nutritional intake. In: McCammon RW, ed. Human growth and development. Springfield, Ill, USA: Charles C Thomas, 1970:63–100.
65. Sontag LW, Munson P, Huff E. Effects on the fetus of hypervitaminosis D and calcium and phosphorus deficiency during pregnancy. Am J Dis Child 1936;51: 302–10.
66. Pao EM, Himes JH, Roche AF. Milk intakes and feeding patterns of breast-fed infants. J Am Diet Assoc 1980;77:540–5.
67. Valadian I, Stuart HC, Reed RB. Studies of illnesses of children followed from birth to eighteen years. Monogr Soc Res Child Dev 1961;26(3):1–125.
68. Garn SM. The genetics of normal human growth. In: Gedda L, ed. De Genetica Medica. Rome: Instituto Gregorio Mendel, 1961:413–32.
69. Roche AF, Wainer H, Thissen D. Predicting adult stature for individuals. Monogr Paediatr 1975;3:1–114.
70. Himes JH, Roche AF, Thissen D. Parent-specific adjustments for assessment of recumbent length and stature. Monogr Paediatr 1981;13:1–88.

71. Lee M, Czerwinski SA, Choh AC, Towne B, Demerath EW, Chumlea WC, Sun SS, Siervogel RM. Heritability of calcaneal quantitative ultrasound measures in healthy adults from the Fels Longitudinal Study. Bone 2004;35:1157–63.
72. Tanner JM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. I. Arch Dis Child 1966;41:454–71.
73. Himes JH. Purposeful assessment of nutritional status. In: Johnston FE, ed. Nutritional anthropology. New York: Alan R. Liss, 1987:85–100.

Interpopulation variation in height among children 7 to 18 years of age

Jere D. Haas and Fabricio Campirano

Abstract

The objective of this review is to examine the degree of variation that exists in the achieved height of preadolescent and adolescent children across populations experiencing favorable conditions that support linear growth. Fifty-three population groups were identified that reported mean heights for economically privileged populations from all major continents. Graphic representation of the heights for these populations indicates that the mean height of preadolescent children differs by 3 to 5 cm, whereas population means begin to diverge from the National Center for Health Statistics/World Health Organization (NCHS/WHO) reference at puberty, with most non-European populations falling to approximately 5 cm below the reference and northern European populations exceeding the reference by a similar amount. We conclude that the evidence for limited interpopulation variation in the height of preadolescents supports consideration of a single growth reference for children up to puberty, but the uncertainty of the causes of the divergence in achieved height during puberty requires further research in order to establish an appropriate adolescent growth reference.

Key words: Adolescence, growth reference, preadolescence, school-aged children, stature

Introduction

In order for a growth reference standard to be useful, it must recognize the variation in achieved growth that exists among and between the populations it is intended to serve. The reference standards in common use today to assess growth of preschool-aged children assume that all populations, regardless of genetic or ethnic background, would achieve similar mean heights if they were afforded the opportunity for most children in that population to achieve a genetically prescribed growth potential. The evidence supporting this assumption is derived from research on interpopulation variation in the height of 7-year-old children reported by Habicht et al. [1] and elaborated by Martorell and Habicht [2]. The basis of this assumption has been challenged by Eveleth and Tanner [3], raising questions not only about the universality of preschool reference standards, but also about the development of useful reference standards for older children and adolescents.

Ulijaszek [4, 5] concluded, after evaluating the ethnic differences in mean heights of 7-year-old children from a large number of studies, that a single reference standard is probably suitable for preadolescents, except perhaps for Asian populations, where heights are about 1.0 to 1.7 cm less than in other population groups from Europe, Africa, North America, and Latin America. This comprehensive study of the published literature also evaluated selected indicators of adolescent growth for ethnic differences and concluded that Asian populations also differed from all others in achieving an earlier age of peak height velocity (PHV), whereas the amount of growth during the year of peak height growth was similar across the various ethnic population groups. Ulijaszek did not examine population differences in achieved size at various ages during adolescence or at attainment of adult size. In the current review of the published literature on attained height of children and adolescents between 7 and 18 years of age, we provide additional evidence for determining whether new growth references for this age group need to be population specific or whether a single reference may be developed that serves all populations.

The authors are affiliated with the Division of Nutritional Sciences, Cornell University, Ithaca, New York, USA.

Please direct queries to the corresponding author: Jere D. Haas, Division of Nutritional Sciences, 220 Savage Hall, Cornell University, Ithaca, NY 14853, USA; e-mail: jdh12@cornell.edu.

Methods

This paper reviews the published research literature on worldwide variation in childhood and adolescent growth. In reviewing the literature that was most relevant to addressing the question of universal versus local growth reference standards, specific criteria were established for selecting studies. The study had to include children and adolescents between 7 and 18 years of age, separated by sex, and mean heights presented in annual age groups. Primary consideration was given to studies of nominally healthy, "well-off," "higher socioeconomic status," or "privileged" subjects who would be exposed to environments that promote maximum linear growth or the greatest probability of achieving individual genetic potential for linear growth. When the data were available, we used the culmination of a positive population secular trend to confirm that this maximal growth potential was achieved. We also sought studies of multiracial populations and international migrants who moved to more advantaged environments. We did not include nationally representative survey data from less-developed countries unless data were presented that allowed high socioeconomic status to be identified as a distinct subgroup.

The data were obtained from a variety of sources. A primary reference was Eveleth and Tanner, *Worldwide Variation in Human Growth*, second edition [3], which included a large number of published and previously unpublished studies available before 1988. This source was supplemented with more recent studies identified through the use of various search engines available over the Internet.* The studies identified by this process are probably representative of the range of populations studied for growth, but they are not likely to be comprehensive or a random selection of all studies conducted on qualified populations.

The data extracted from these studies include height for annual age groups expressed as means or medians, with standard deviations or percentiles when available; age expressed as age ranges, means, or medians for annual age groups; demographic and social characteristics such as sex, socioeconomic and health conditions, racial or ethnic classification, and migration status. The vast majority of studies employed cross-sectional sampling. The few longitudinal studies were included if they represented the higher socioeconomic class of the particular country. Most of the data were obtained directly from published tables. Several studies reported data only in graphs, from which estimated values were extracted by age group. Percentile data were converted to the 10th and 90th percentiles. All ages were centered at the mean of the age group, which tended to be at the half year of one-year age groups. So that evaluation could be made of age trends within a population, only studies that reported yearly mean heights over at least a 3-year period were included.

The data are expressed in graphs according to the major geographic areas where ancestral populations evolved and lived prior to the colonial period (ca 1492 AD for Africa, Asia, and Europe), as well as the areas of major population expansion from ancestral homes after 1492 (Europe, the Americas, and Australia). Although this classification of populations resembles the ethnic classification used by Ulijaszek [5], it differs in its emphasis on geographic populations rather than ethnic groupings. No simple classification can identify ancestral subpopulations that reflect reproductive isolation or other evolutionary processes that may have led to genetic differences in achieved growth. The classification employed in the current review is an attempt to group populations by broad, geographically based regions of ancestral origins. We recognize that these groupings do not adequately reflect the isolation of precolonial-period populations because of the considerable admixing of populations within geographic regions before the colonial period and more recently between populations. However, the groupings do represent major geographic divisions that facilitate interpopulation comparisons around the world and reflect the current population diversity for which a contemporary growth reference will be used.

Data are presented graphically for selected ages between 7 and 18 years. No statistical analysis was performed to test for the significance of population differences in achieved growth or age trends in growth.

Results

The results are presented in four sections. The first section presents the background on the studies identified for this analysis. This is followed by presentation of interpopulation comparisons of mean heights of boys and girls within four geographic regions at each of four ages: 7, 10, 13, and 17 years. Age trends in the deviation in mean achieved height from the National Center for Health Statistics/World Health Organization (NCHS/WHO) reference [6] are presented for selected samples representing the tallest or most privileged population in each geographic area and, where appropriate, according to international migration status. To further evaluate pubertal growth and its relation to prepubertal growth, we estimated the approximate increment in growth during the pubertal period and compared it with the mean height of the population at the beginning of puberty.

* These included PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi), High Wire Press (http://highwire.stanford.edu/), OAIster (http://oaister.umdl.umich.edu), the Cornell University Library Catalog (http://catalog.library.cornell.edu), and several sources obtained from a network of professional contacts.

Description of the studies

A total of 53 study samples of nominally healthy children and adolescents representing four major ancestral population regions were identified for the analysis of interpopulation variation in height growth. The study

TABLE 1. Population samples included in this review

Reference	Region of origin and country	Place of residence or population group	Author and publication year	Study years
	Africa			
7	Morocco	Amsterdam, Netherlands	Fredriks et al., 2004	1997
8	Nigeria	Ibadan	Janes, 1970[a]	1962–70
9	Kenya	Nairobi	Kulin et al., 1982[a]	1980
10	Cuba	Havana	Vidallet et al., 2003	n.d.
11	Jamaica	Kingston	Ashcroft et al., 1964[a]	n.d.
12	African-American	USA—national	NCHS[a], n.d.	1976-80
13	African-American	USA—Texas	Shutte, 1980[a]	n.d.
14	African-American	USA—national	Karpati et al., 2002	1988–94
	East Asia			
15	China	Beijing	Li et al., 1999	1995
16	China	Hong Kong	Fung et al., 1985[a]	1984
17	Taiwan	National	Chen et al., 2003	1993–96
18	Japan	National	Kikuta et al., 1987[a]	1980
	Japan	National	Sano et al., 2006[b]	1984–2001
19	Thailand	Bangkok	Khanjanasthti et al., n.d.[a]	1981–84
20	Native American	Mexican-Americans	Ryan et al., 1999	1988–94
21	Native American	Plains Amerindian	Zephier et al., 2006	2002–03
22	Native American	Chippewa	Johnston et al., 1979[a]	n.d.
	South and West Asia			
23	Pakistan	Karachi	Hakeem et al., 2004	2000
24	Pakistan	Birmingham, UK	Kelly et al., 1997	1989
25	Pakistan	UK	Marshall et al., n.d.[a]	1982–83
26	India	Calcutta	de Onis et al., 2001	1982
27	Iran	Isfahan	Aminorroaya et al., 2003	1997
28	Turkey	Istanbul	Neyzi et al., 1973[a]	n.d.
29	Turkey	Amsterdam, Netherlands	Fredriks et al., 2003	1997
30	Jordan	Jordan	Hasan et al., 2001	1997
31	United Arab Emirates	National	Al-Hourani et al., 2003	1998–99
32	Saudi Arabia	National	Al-Nuaim et al., 1996	1994–95
	Europe			
33	Sweden	Stockholm	Lindgren, 1976[a]	1964–73
34	Sweden	Gothenburg	Wikland et al., 2002	1992
35	Germany	Jena	Hesse, 1988[a]	1979–87
36	Norway	Oslo	Brundtland et al., 1980[a]	1975
37	Netherlands	Amsterdam and others	Fredriks et al., 2000	1996–97
38	Belgium	National	Vercauteren et al., 1985[a]	1980–82
39	Ireland	Dublin	Hoey et al., 1987[a]	n.d.
40	UK	England	Cole et al., 2006	1978–90
41	UK	England	Rona et al., 1986[a]	1982–83
42	Czechoslovakia	National	Blaha, 1986[a]	1985
43	Spain	National	Hernandez et al., 1988	n.d.
44	Italy	National	Cacciari et al., 2002	1996–2002
45	Italy	Sardinia	Sanna and Soro, 2000	1996
46	Greece	National	Mantzagriotou-Meimarides, n.d.[a]	1983–84

continued

samples, with selected characteristics, are listed in **table 1**. These include 8 samples (15%) from populations of African origin, 9 (17%) from populations of East Asian origin, 10 (19%) from populations of South or West Asian origin, and 26 (49%) from populations of European origin. Twenty-four samples were taken from migrant populations, including 4 from migrants to Europe, 10 from migrants to North America, 7 from migrants to South America or the Caribbean, and 1 from migrants to Australia. Data for 23 (43%) of the population samples were obtained from Eveleth and Tanner [3], and data for the remaining 30 samples were obtained from other published sources. For consistency and ease of reporting, all results are presented with the NCHS/WHO reference median [6] derived from the growth reference originally published by the US National Center for Health Statistics (NCHS) [57, 59] and distributed by the US Centers for Disease Control and Prevention (CDC) in collaboration with the World Health Organization (WHO) [6]. It represents data from several national probability samples of US youths and adolescents 6 to 18 years of age, collected during two rounds of the National Health Surveys of 1963–1970.

Height variation across geographic regions

Figures 1 and **2** present the variation in mean heights within each geographic region at four selected ages. At 7 years of age, the maximum mean heights for boys (**fig. 1**) in each region range from 122 cm for boys from Nigeria (Africa) to 127 cm for Northern Plains American Indians (ancestors from central Asia). These values are comparable to the median height of 124 cm for the NCHS/WHO reference population, and very little difference is seen between the tallest boys in Asia and Europe. Seven-year-old girls (**fig. 2**) follow the same trend as boys; American Indian girls as a group are slightly taller than girls from Europe and Africa, and the mean heights of girls in the tallest samples from each region are within 3 cm of the reference median of 123 cm. The range between the shortest and tallest mean or median values within a geographic region is approximately 9 cm. The midpoint across population means within a region is about 1 cm below the NCHS/WHO reference median for both sexes.

At 10 years of age, there are similar ranges in mean heights across regions for both boys and girls, and the midpoints of the heights are very similar to the NCHS/WHO reference median values of 140 and 141 cm for boys and girls, respectively. The tallest boys in each region do not differ greatly, with heights between 140 and 144 cm. For 10-year-old girls, the maximum heights in each region vary from 140 cm for African-American girls to 144 cm for girls in urban Mexico.

At 13 years of age, the mean heights of boys vary by a similar amount across regions and generally average about 2 cm below the NCHS/WHO reference median. The tallest boys in each region range from 159 to 163 cm. The mean heights of 13-year-old girls in Africa and Asia are uniformly about 4 cm less than the values in Europe and the NCHS/WHO reference value. The mean heights of the tallest girls in each region range from 158 to 161 cm, while the NCHS/WHO reference median is 159 cm.

By 17 years of age, the greatest differences are observed across regions, with northern European boys and girls being the tallest and Asian boys and girls the shortest. The variation among samples within regions,

TABLE 1. Population samples included in this review *(continued)*

Reference	Region of origin and country	Place of residence or population group	Author and publication year	Study years
47	Cyprus	Cyprus	Savva et al., 2001	1999–2000
	European migrants to:			
48	Chile	Santiago	Yulton et al., 1990	1983
49	Venezuela	National	Mendez et al., 1993	n.d.
50	Venezuela	National	Landaeta-Jimenez et al., 2002	1995–96
51	Mexico	Mexico City	Faulhaber, 1989[a]	1977–80
52	Mexico	Urban Veracruz	Brewis, 2003	2001
53	Argentina	National	Lejarraga, 1986[a]	1974–75
54	Brazil	Santo Andre	Marques et al., 1982[a]	1978
55	Australia	National	Pyke, 1986[a]	1985
56	Canada	National	Nutrition Canada, 1980	1970–72
57	USA	National	Hamill et al., 1977	1962–74
58	USA	National	CDC 2000	1977–2000

n.d., year of study not stated; NCHS, National Center for Health Statistics; CDC, Centers for Disease Control and Prevention
a. Data from tables in Eveleth and Tanner [3].
b. Unpublished data reported by S. Yamamoto at the conference on the Development of an International Growth Standard for Preadolescent and Adolescent Children, in Geneva, 16–19 January 2006.

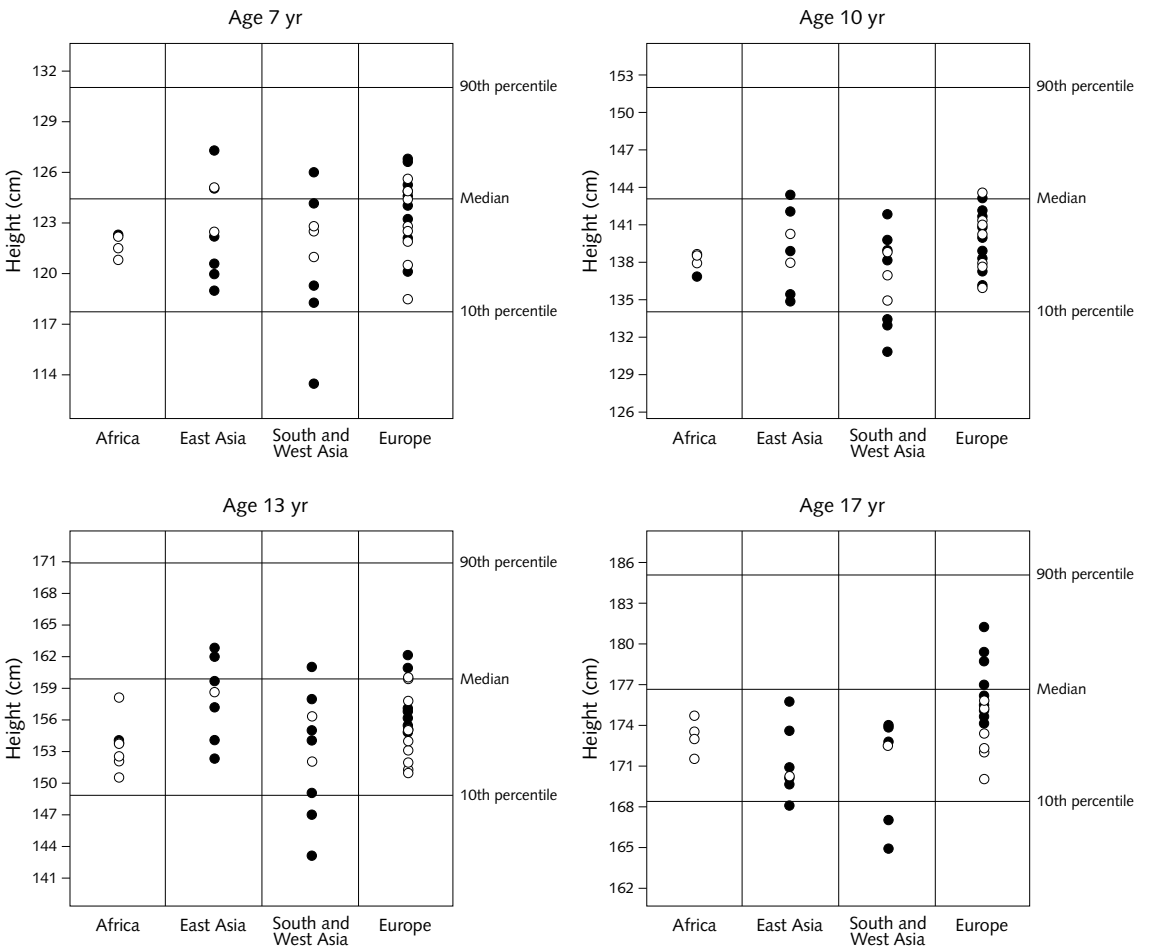

FIG. 1. Mean heights of "privileged" boys at selected ages according to geographic region and migration status. Closed circles, natives; open circles, migrants. Median, 10th, and 90th percentiles from NCHS/WHO (1983) [6]

however, is relatively small. The tallest boys are 174 to 175 cm in height in all of the regions except Europe (the Netherlands), where the tallest boys have a height of 181 cm and the NCHS/WHO reference median is 177 cm. For 17-year-old girls, the heights of the tallest girls in each region range from 161 cm (China) to 169 cm (the Netherlands), with a NCHS/WHO reference median of 163 cm.

Selected samples from each region

If a reference is to be based on the healthiest children in a population, and the healthiest children are most likely to achieve their genetically programmed potential for linear growth, then comparison of the tallest sample of children within an ethnic or geographic group may provide information on genetic differences in height. The comparison of maximum achieved height across regions would be useful to justify decisions on whether population-specific references are appropriate for certain age groups.

Figures 3 and **4** present age trends in height-for-age z-scores of selected samples of the tallest boys and girls in each of the four geographic regions, plus a few selected ethnic and migrant groups. East Asia is represented by urban China and Japan, South and West Asia are represented by Turkey, Africa is represented by African-Americans in the United States, South and Central Europe are represented by Italy, North Europe is represented by the Netherlands, and Native Americans are represented by Mexican-Americans and Plains American Indians in the United States. The populations selected for these figures may be considered as representing major geographic populations where socioeconomic conditions appear most favorable to support maximum linear growth.

The z-scores for the mean heights of boys (**fig. 3**) from these populations track along the median (z-score = 0) of the NCHS/WHO reference, with an average difference between the shortest and tallest groups of about 1.0 SD or approximately 6 cm, between the ages of 7 and 13 years. By the age of 15 years, the heights of

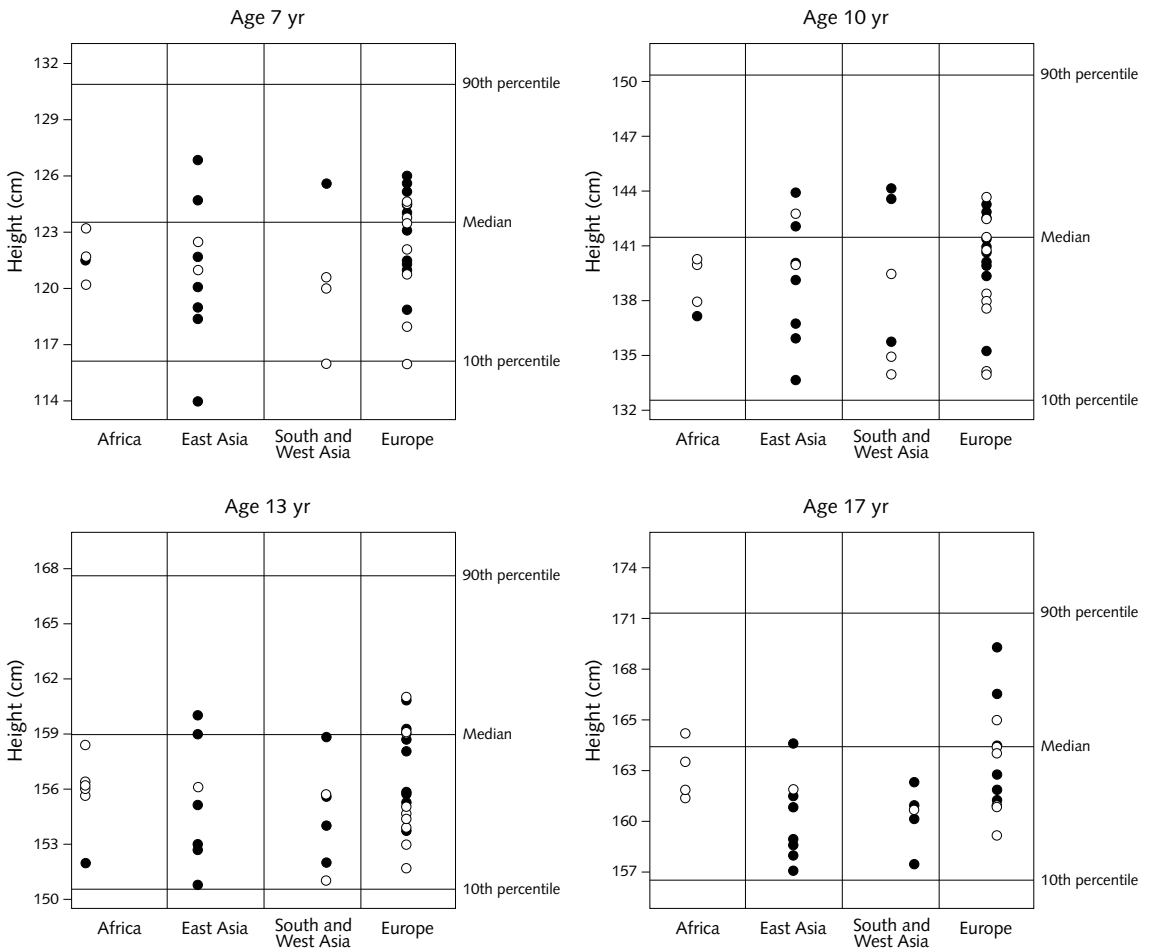

FIG. 2. Mean heights of "privileged" girls at selected ages according to geographic region and migration status. Closed circles, natives; open circles, migrants. Median, 10th, and 90th percentiles from NCHS/WHO (1983) [6]

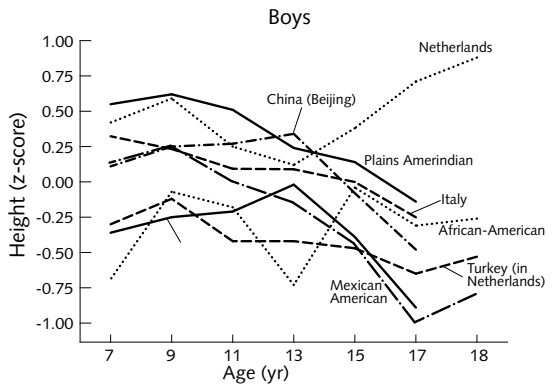

FIG. 3. Mean height-for-age z-scores for preadolescent and adolescent boys from selected "privileged" populations. Z-scores calculated from NCHS/WHO (1983) [6]

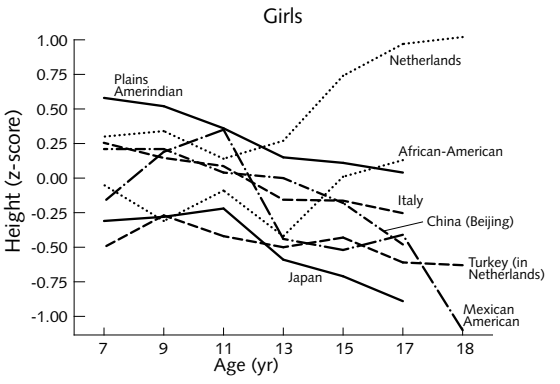

FIG. 4. Mean height-for-age z-scores for preadolescent and adolescent girls from selected "privileged" populations. Z-scores calculated from NCHS/WHO (1983) [6]

children in several groups (Italian, Plains Indian, Chinese, Japanese, and Mexican-American) have declined relative to the NCHS/WHO reference. In contrast, the sample from the Netherlands begins to diverge above the reference and all other groups. At 17 years of age, the Dutch boys are 5 cm taller than the reference and the Mexican-American and Japanese boys are approximately 5 cm shorter than the reference, a difference between the extremes of nearly 2 SD.

The trend for girls (**fig. 4**) is similar to that for boys. Consistent with the earlier timing of the pubertal growth spurt, the mean values for the Dutch, Mexican-American, and Japanese girls begin to diverge from the reference approximately 2 years earlier than observed for boys. By 17 years of age, the Dutch girls are 7 cm taller than the NCHS/WHO reference and the Mexican-American and Japanese girls are approximately 5 cm below the reference.

Height growth during puberty

Considering that the greatest interpopulation differences in achieved height appear to occur during puberty, we examined the amount of growth that occurs for each population between the approximate initiation of puberty and the approximate achievement of adult height. **Figures 5** and **6** show the interpopulation variation according to geographic region for the calculated difference in mean heights from 11 to 17 years in boys and from 10 to 17 years for girls. European populations, in general, have the greatest amount of growth during puberty and East Asian populations the least. The ranges within a region are very large, especially for East Asia and Europe, where there are more populations represented. The wide range represented by European populations seems to be divided between those with a higher approximate growth rate from Northern Europe and those with the lowest rate from Southern Europe and among migrants from Europe to less developed regions, with values computed from the NCHS/WHO reference in the middle to lower end of this range.

In order to investigate whether the pubertal period could be a period of catch-up for slower prepubertal growth, we plotted the pubertal growth shown in **figures 5** and **6** by the height at the beginning of the pubertal period. **Figures 7** and **8** present these relationships. In general, children from populations that enter puberty with the lowest stature appear to have the greatest growth in stature during puberty, with significant ($p < .05$) Pearson correlations of –0.32 and –0.40 for boys and girls, respectively. Within geographic subgroups, the negative association between initial stature and pubertal change in stature persists and is generally stronger than it is for the entire sample. The strongest associations are seen for the eight populations from central and Southern Europe ($r = -0.91$ and –0.86 for boys and girls, respectively) and the eight East Asian populations ($r = -0.84$ and –0.81). In South and West Asian populations, there is no correlation for the six samples of boys, but a very strong correlation ($r = -0.96$) for the four samples of girls. The correlations for the seven populations of migrants from Europe are –0.74 and –0.53 for boys and girls, respectively. The trends for the African and Northern European regions, which are each represented by only three populations, are difficult to interpret.

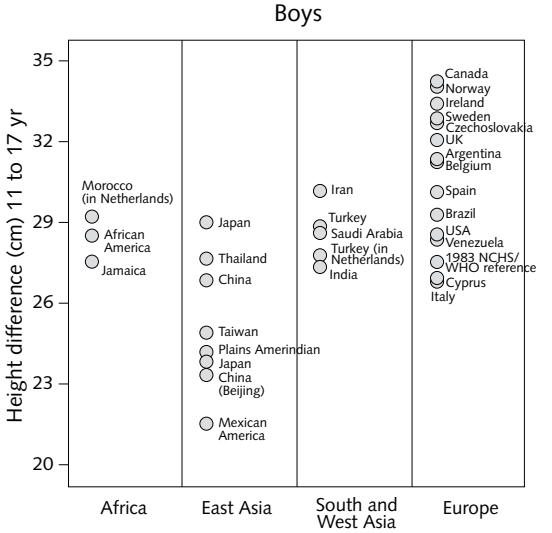

FIG. 5. Approximate height growth between 11 and 17 years for selected samples of boys from four regions

FIG. 6. Approximate height growth between 10 and 17 years for selected samples of girls from four regions

Discussion

The data on achieved height of 7-year-old children from more recent studies identified for this analysis confirm previous observations by Habicht et al. [1] and Ulijaszek [5] that there are very few interpopulation differences in average height growth of preschool children from the highest socioeconomic levels. These authors therefore conclude that genetic factors have a minimal impact on the interpopulation variation in early linear growth. The exceptions to this trend are the East Asian populations, which tend to be about 2 cm shorter than populations from other regions. However, as Ulijaszek [5] points out, many of these Asian populations appear to be experiencing a secular trend in height growth, which when completed may close the gap for attained height during the prepubertal period.

The trend for limited interpopulation differences in attained height appears to continue until the onset of puberty. The current study does not attempt to systematically analyze the differences in height between socioeconomic groups within the same ethnic or geographic

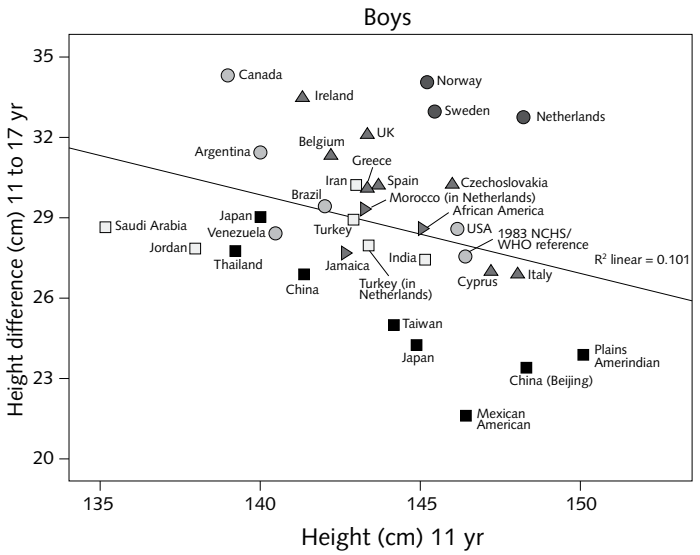

FIG. 7. Approximate pubertal height growth in relation to prepubertal height at 11 years in selected samples of boys from four regions. Line represents least square regression

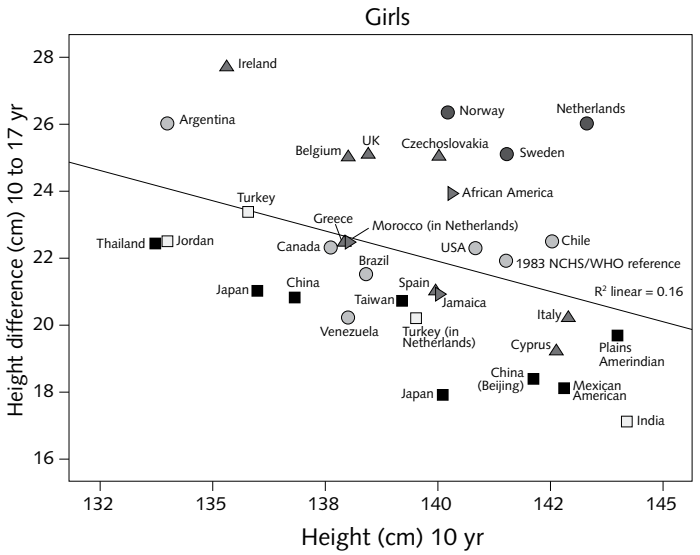

FIG. 8. Approximate pubertal height growth in relation to prepubertal height at 10 years in selected samples of girls from four regions. Line represents least square regression

population. However, according to the limited data that are available from studies reviewed here, the differences in mean height due to socioeconomic variation within a population are generally greater than the differences in means of the most privileged children across ethnic or geographic groups. This observation was previously reported by Habicht et al. [1] and by Martorell and Habicht [2] for 7-year-old children, and it continues to be valid for children through at least 10 years of age in the current review.

Of those nominally healthy samples that differ from the NCHS/WHO reference, most diverge at the approximate age when the pubertal growth spurt begins, and by the end of puberty the differences are greater than at the outset of puberty. It is difficult to determine whether the apparent reduction in linear growth rate during puberty for most of these groups is due to genetic factors or sampling errors resulting from age cohort differences in previous environmental exposures. Most of the studies reported here are from cross-sectional samples, where one might suspect that, for some populations, the stage of a positive secular trend might be experienced differently by different age cohorts. Younger children may have been exposed to more favorable environments during the critical preschool age years than were the older children, who also experienced limited catch-up growth. The observation that populations that are shortest when entering puberty appear to grow the most during puberty may suggest that some catch-up potential exists at this rather late but dynamic phase of linear growth. However, sorting out population secular trend effects from individual catch-up growth is impossible with cross-sectional data and can only be partially resolved with longitudinal data that are currently available in the published literature. Thoughtful examination of longitudinal data for individual children from sequential studies of the same population groups over several decades would help resolve this issue.

The apparent faltering of growth for most non-European populations seems to contradict the findings of Ulijaszek [4, 5], who examined two features of the pubertal growth spurt reported in longitudinal studies of a variety of populations. He showed that the amount of linear growth occurring during the year of PHV did not differ across a wide range of the world's population groups. Further, the age at which PHV was observed is very similar across populations, with the exception of some Asian populations that began their peak growth earlier than the other groups. Exactly how these observations influence our interpretation of interpopulation variation in achieved height throughout puberty needs further investigation. The timing and intensity of the adolescent growth spurt are only two parameters that influence achieved size. It is also necessary to examine the amount of growth that occurs during the years before and after the year of PHV as well as the length of the pubertal growth period in order to determine the total amount of pubertal growth that is added to the prepubertal achieved height that determines final adult stature.

The genetic and/or environmental factors that influence the onset and duration of puberty within and between populations preclude speculation about factors that affect variation in height across populations. It might be useful to analyze population differences in achieved height relative to developmental landmarks of puberty, such as those observed from secondary sex characteristics. This would allow interpopulation comparisons of height at specific chronological ages for population groups that share similar timing and tempo of pubertal development. Where longitudinal data exist, it would be informative to extend the analysis performed by Ulijaszek [5] to include population comparisons of the growth experience throughout puberty, by relating the timing and amount of growth during the year of PHV with stature achieved in early adulthood.

Conclusions about whether interpopulation variation in achieved growth may be due to ethnic or genetic differences are tempered by the paucity of recent data from several important population groups, including Africans in Africa and indigenous populations in the Americas. North America, South America, and Africa are large continents with considerable genetic, ethnic, and environmental diversity among the indigenous populations that have resided there for thousands of years. The available evidence obtained for this review is limited to a few older studies on populations that may have experienced favorable environmental conditions for achieving linear growth potential. These populations were probably still experiencing a secular trend in linear growth or were represented by migrant or native populations with substantial genetic admixture from colonizing populations.

The greater achieved stature of the Northern European populations compared with the NCHS/WHO reference and all other populations requires further study. Whether these populations represent a growth potential that is achievable by all populations living under ideal environmental conditions depends on whether one can subscribe to the position that "bigger is better" with regard to linear growth. It is yet to be determined whether environmental and dietary factors unique to Northern Europe or genetic isolation and homogeneity may have accounted for this unusual growth pattern. A study of Northern European migrants to favorable environments in a variety of developed countries could be useful in explaining the very tall stature of this population group.

A fundamental problem with any exercise that attempts to describe human biological variation as a racial phenomenon is that the race concept as applied to humans is invalid. Although race is used as a social,

demographic, and political instrument, it has no legitimacy in human biology [60]. The current analysis uses the term *geographic* or *ethnic* group to identify subpopulations that may show sufficient reproductive isolation and/or differences in ancestral environments that could have affected genotypic variation. Whether the degree of isolation and ancestral environmental exposures experienced by these subpopulations is sufficient to affect the genes that control linear growth has not been tested. However, if subpopulation differences in achieved growth are observed under environmental conditions that support maximum linear growth, there is sufficient justification to explore genetic explanations for these differences.

Conclusions

This review of worldwide variation in preadolescent and adolescent growth indicates that some differences in achieved linear growth exist across a broad array of human populations that could be considered nominally healthy. However, these patterns are not uniform across all ages. There is sufficient evidence to suggest that average linear growth up to the onset of puberty is similar across populations that experience favorable growing environments. Potential population differences in the initiation and progression of puberty may account for the large differences seen in achieved height during and after puberty. However, one cannot discount the effect of secular trends that confound age-specific growth patterns across cohorts in explaining some of the divergence of achieved heights with increasing age through puberty. The large differences between the heights of healthy young adults in Japan and Northern Europe compared with heights in the rest of the world's populations suggest very different growth patterns during puberty and the possibility that unique environments and genetic factors may account for these growth differences.

References

1. Habicht JP, Yarbrough C, Martorell R, Malina R, Klein RE. Height and weight standards for preschool children. How relevant are ethnic differences in growth potential? Lancet 1974;1(7858):611–5.
2. Martorell R, Habicht JP. Growth in early childhood in developing countries. In: Falkner F, Tanner JM, eds. Human growth: a comprehensive treatise. Vol 3. 2nd ed. New York: Plenum, 1985:241–62.
3. Eveleth PB, Tanner JM. Worldwide variation in human growth. 2nd ed. Cambridge: Cambridge University Press, 1990.
4. Ulijaszek S. Between-population variation in pre-adolescent growth. Eur J Clin Nutr 1994;48(suppl 1):S5–14.
5. Ulijaszek S. Ethnic differences in patterns of human growth in stature. In: Martorell R, Haschke F, eds. Nutrition and growth. Nestle Nutr Workshop Ser Pediatr Program 2001;47:1–15.
6. World Health Organization. Measuring change in nutritional status. Guidelines for assessing the nutritional impact of supplementary feeding programs for vulnerable groups. Geneva: WHO, 1983.
7. Fredriks M, van Buuren S, Jeurissen SE, Dekker FW, Verloove-Vanhorick SP, Wit JM. Height, weight, body mass index and pubertal development references for children of Moroccan origin in the Netherlands. Acta Paediatr 2004;93:817–24.
8. Janes MD. The effect of social class on the physical growth of Nigerian Yoruba children. Bull Int Epidemiol Assoc 1970;20:127–36. (cited in Eveleth and Tanner [3]).
9. Kulin HE, Bwibo N, Mutie D, Santner SJ. The affect of chronic childhood malnutrition on pubertal growth and development. Am J Clin Nutr 1982;36:527–36.
10. Vidallet E, Rodríguez G, Carnot J, Perez A, Duane OJ. Indicadores antropométricos en la evaluación nutricional en adolescentes del sexo masculino. Rev Cubana Pediatr 2003, 75(2). Available at: http://www.scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0034-75312003000200001&lng=es&nrm=iso. Accessed 6 September 2006.
11. Ashcroft MT, Heneage P, Lovell HG. Heights and weights of Jamaican schoolchildren of various ethnic groups. Am J Phys Anthropol 1966;24:35–44.
12. National Center for Health Statistics. Unpublished data, n.d. (cited in Eveleth and Tanner [3]).
13. Shutte J. Growth of black male adolescents. Hum Biol 1980;52:193–204.
14. Karpati AM, Rubin CH, Kieszak SM, Marcus M, Troiano RP. Stature and pubertal stage assessment in American boys: the 1988–1994 Third National Health and Nutrition Examination Survey. J Adolesc Health 2002;30;205–12.
15. Li H, Leung SS, Lam PK, Zhang X, Chen XX, Wang SL. Height and weight percentile curves of Beijing children and adolescents 0–18 years, 1995. Ann Hum Biol 1999; 26:457–71.
16. Fung KP, Lan SP, Chow OKW, Baber F, Chu SY, Tsoi NS, Lun KW, Chan SC, Lam TK. A survey of growth of Hong Kong children. Hong Kong J Paediatr 1985;2:105–16 (cited in Eveleth and Tanner [3]).
17. Chen JY, Chang HY, Pan WH. A modified locally weighted method for developing reference standards for height, weight, and body mass index of boys and girls aged 4 to 18 in Taiwan. Hum Biol 2003;75:749–70.
18. Kikuta F, Takaishi M. Studies on physical growth standards for schoolchildren in Japan. Part I. Centile curves for height and weight based on cross-sectional data and consideration of secular trend of the centile curves. Jpn J Child Health 1987;46:27–33 (cited in Eveleth and Tanner [3]).
19. Khanjanasthti P, Chawewan C, Siripat W. Growth of Bangkok children, 0–18 years. Bangkok, Mahidol University, n.d. (cited in Eveleth and Tanner [3]).

20. Ryan AS, Roche AF, Kuczmarski RJ. Weight, stature, and body mass index data for Mexican Americans from the Third National Health and Nutrition Examination Survey (NHANES III, 1988–1994). Am J Hum Biol 1999; 11:673–86.
21. Zephier E, Himes JH, Story M, Zhou X. Increasing prevalences of overweight and obesity in Northern Plains American Indian children. Arch Pediatr Adolesc Med 2006; 160:34–9.
22. Johnston F, Schell L. Anthropometric variation of Native American children and adults. In: Laughlin WS, Harper AB, eds. The first American: origins, affinities and adaptations. New York: Gustar Fischer, 1979:275–91.
23. Hakeem R, Shaikh AH, Asar F. Assessment of linear growth of affluent urban Pakistani adolescents according to CDC 2000 references. Ann Hum Biol 2004;31:282–91.
24. Kelly AM, Shaw NJ, Thomas AM, Pynsent PB, Baker DJ. Growth of Pakistani children in relation to the 1990 growth standards. Arch Dis Child 1997;77:401–5.
25. Marshall BM, Marshall WA, Nicoll AG, Peters J, Ulijaszek SJ. Asian heights and weights compared with standard centiles. Unpublished manuscript, n.d. (cited in Eveleth and Tanner [3]).
26. de Onis M, Dasgupta P, Saha S, Sengupta D, Blossner M. The National Center for Health Statistics reference and the growth of Indian adolescent boys. Am J Clin Nutr 2001;74:248–53.
27. Aminorroaya A, Amini M, Naghdi H, Zadeh AH. Growth charts of heights and weights of male children and adolescents of Isfahan, Iran. J Health Popul Nutr 2003; 21:341–6.
28. Neyzi O, Alp H, Yalcindag A. Heights and weights in Turkish children. Environ Child Health 1973;19(2A):5–13 (cited in Eveleth and Tanner [3]).
29. Fredriks AM, van Buuren S, Jeurissen SE, Dekker FW, Verloove-Vanhorick SP, Wit JM. Height, weight, body mass index and pubertal development reference values for children of Turkish origin in the Netherlands. Eur J Clin Nutr 2003;162:788–93.
30. Hasan MA, Batieha A, Jadou H, Khawaldeh AK, Aijlouni K. Growth status of Jordanian schoolchildren in military-funded schools. Eur J Clin Nutr 2001;55:380–6.
31. Al-Hourani HM, Henry CJ, Lightowler HJ. Prevalence of overweight among adolescent females in the United Arab Emirates. Am J Hum Biol 2003;15:758–64.
32. Al-Nuaim AR, Bamgboye EA, Al-Herbish A. The pattern of growth and obesity in Saudi Arabian male school children. Int J Obes Relat Metab Disord 1996;20:1000–5.
33. Lindgren G. Height, weight and menarche in Swedish urban school children in relation to socio-economic and regional factors. Ann Hum Biol 1976;3:501–28.
34. Wikland KA, Luo ZC, Niklasson A, Karlberg J. Swedish population-based longitudinal reference values from birth to 18 years of age for height, weight and head circumference. Acta Paediatr 2002;91:739–54.
35. Hesse V. Unpublished growth data from Jena, 1988 (cited in Eveleth and Tanner [3]).
36. Brundtland GH, Liestol K, Walloe LW. Height, weight and menarcheal age of Oslo schoolchildren during the last 60 years. Ann Hum Biol 1980;7:307–22.
37. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in the Netherlands 1955–1997. Pediatr Res 2000; 47:316–23.
38. Vercauteren M, Susanne C. The secular trend of height and menarche in Belgium: Are there any signs of a future stop? Eur J Pediatr 1985;144:306–9.
39. Hoey H, Cox L. Irish standards for triceps and subscapular skinfold thickness. Ir Med J 1987;80:312–5.
40. Cole TJ, Pan H. LMS growth, an Excel addin module for converting growth data to z-scores. Available at: http://homepage.mac.com/tjcole/FileSharing1.html. Accessed 6 September 2006.
41. Rona RJ, Chinn S. National Study of Health and Growth: social and biological factors associated with height of children from ethnic groups living in England. Ann Hum Biol 1986;13:453–71.
42. Blaha P. Anthropometric studies of the Czechoslovak population: from 6 to 55 years (2 volumes). Czechoslovak Spartakiade, 1985. Prague, 1986 (cited in Eveleth and Tanner [3]).
43. Hernandez M, Castellet J, Narvaiza JL, Rincon JM, Ruiz I, Sanchez E, Sobradillo B, Zurimendi A. Curves y tablas de crecimiento. Instituto de Investigación sobre Crecimiento y Desarrollo. Madrid: Editorial Garsi. 1988.
44. Cacciari E, Milani S, Balsamo A, Dammacco F, De Luca F, Chiarelli F, Pasquino AM, Tonini G, Vanelli M. Italian cross-sectional growth charts for height, weight and BMI (6–20y). Eur J Clin Nutr 2002;56:171–80.
45. Sanna E, Soro MR. Anthropometric changes in urban Sardinian children 7 to 10 years between 1975–1976 and 1996. Am J Hum Biol 2000;12:782–91.
46. Mantzagriotou-Meimarides M. Unpublished Greek national growth curves, n.d. (cited in Eveleth and Tanner [3]).
47. Savva SC, Kourides Y, Tornaritis M, Epiphaniou-Savva M, Tafouna P, Kafatos A. Reference growth curves for Cypriot children 6 to 17 years of age. Obes Res 2001;9:754–62.
48. Youlton R, Valenzuela C. Growth patterns in height and weight in children aged 0 to 17 years and cranial circumference in children aged 0 to 2 years from medium-high and high socioeconomic status in Santiago. Comparison with growth in children from medium-low and low status in the northern area of Santiago. Rev Chil Pediatr 1990;spec no.:1-22 (in Spanish).
49. Mendez HM, Landaeta-Jimenez M, Saab L. Curvas y tablas de crecimiento. Caracas, Venezuela: FUNDACRESA, 1993.
50. Landaeta-Jimenez M, Perez BM, Escalante Y. Fatness and fat distribution by social stratum in Venezuelan youths (in Spanish). Arch Latinoam Nutr 2002;52:128–36.
51. Faulhaber J. Crecimiento: somatometría de la adolescencia. México City, Mexico: Universidad Nacional Autónoma de México, 1989 (cited in Eveleth and Tanner [3]).
52. Brewis A. Biocultural aspects of obesity in young Mexican schoolchildren. Am J Hum Biol 2003;15:446–60.
53. Lejarraga H. Peso y talla de 15,214 adolescentes de todo el pais. Tendencia secular. Arch Argent Pediatr 1986;84:219–35 (cited in Eveleth and Tanner [3]).
54. Marques RM, Marcondes E, Berguo E, Prendi R, Unes J. Crescimento e desenvolvimento Pubertario en criancas e adolescentes Brasilieros. II. Altura e peso. Ed. Brasileira de Ciencias, Sao Paulo, 1982 (cited in Eveleth and Tanner [3]).

55. Pyke JE. The Australian schools fitness test, The Australian Council for Health, Parkside, S.A., 1986 (cited in Eveleth and Tanner [3]).
56. Nutrition Canada. Anthropometry report—height, weight and body dimensions. Ottawa: Bureau of Nutritional Sciences, Health Protection Branch, Health and Welfare Ministry, 1980 (cited in Gibson, RS. Principles of nutritional assessment. New York: Oxford University Press, 1990).
57. Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF. NCHS growth curves for children birth–18 years, United States. Vital Health Stat 1977;(165):i–iv, 1–74.
58. Centers for Disease Control and Prevention. 2000 CDC Growth Charts: United States. Data tables. Selected percentiles and LMS parameters. Hyattsville, Md, USA: National Center for Health Statistics, 2000. Available at: http://www.cdc.gov/growthcharts. Accessed 3 August 2006.
59. Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF, Moore WM. Physical growth: National Center for Health Statistics percentiles. Am J Clin Nutr 1979;32:607–29.
60. Relethford JH. The human species: an introduction to biological anthropology. 5th ed. Boston, Mass, USA: McGraw-Hill, 2003.

Theoretical considerations related to cutoff points

David Pelletier

Abstract

The cutoff points for creating anthropometric indicators of size and growth can be established by three distinct methods: statistical, risk-based, and prescriptive. The theoretical, philosophical, and technical bases for these are quite distinct, but the implications of each method for applications at population and individual levels can be explored by using a common conceptual model. This model posits that any observed anthropometric distribution is a mixed distribution of two (or more) subpopulations, representing some individuals who are or will remain healthy (the specificity distribution) and those who are or will become unhealthy (the sensitivity distribution). The performance and appropriateness of cutoff points based on statistical, risk-based, and prescriptive criteria depend upon the relative sizes of these two subpopulations in a given context, the distance between their means, and the strength and shape of the relationship between the anthropometric indicator and the health outcomes that define these two subpopulations. The risk-based and prescriptive methods both require substantial epidemiologic evidence if they are to fulfill their theoretical and public health expectations, and both face normative (ethical) trade-offs in establishing cutoff points. The prescriptive method faces even stronger normative challenges, especially in relation to overweight and obesity, because its explicit claim regarding the desirable size and growth of children and adolescents may understate the importance of individuality and overstate the strength of the relationship (and the evidence for the relationship) between size, growth, and future health. These concerns are most pronounced for applications at the individual level and for mild-to-moderate elevations of body-mass index and other indicators.

The author is affiliated with the Division of Nutritional Sciences, Cornell University, Ithaca, New York, USA.

Please direct queries to the author: David Pelletier, Division of Nutritional Sciences, Cornell University, 378 MVR Hall, Ithaca, NY 14853, USA: e-mail: Dlp5@cornell.edu.

Key words: Sensitivity, specificity, norms, cutoff points, tradeoffs

The choice of cutoff points for a nutritional indicator has a profound impact on the performance of the indicator in various clinical and public health applications at the individual and population levels. This is true when international *reference values* for growth are used, as in previous decades, and it applies with even greater force to the case of *growth standards* because of the normative claims made in the latter case. The aims of this paper are to:

» Distinguish the many purposes to which growth data are applied, each of which has its own implications and requirements vis-à-vis choice of cutoff points;
» Review some basic theoretical considerations related to the choice of cutoff points for child growth, in relation to each of the above purposes;
» Highlight some key differences between low-end versus high-end cutoff points as indicators of undernutrition and overnutrition, respectively;
» Develop an integrated framework, based on the above considerations and those from other papers in this series, to guide further analysis and deliberations related to the choice of cutoff points.

The considerations outlined in this paper suggest that several key issues will need to be resolved in future discussions related to the child and adolescent growth project. First, there is a need to discuss whether the practical advantages of using single, uniform cutoff points for all or most applications (as has been the tendency in the past) outweigh the associated performance inefficiencies. Second, in light of the trade-offs related to cutoff points, there is a need to discuss whether the value judgments associated with use of a growth *standard* can be justified in the case of childhood and adolescence (as distinct from infancy) and in relation to overweight and obesity (as distinct from undernutrition). Third, an effort should be made to directly engage some of the end-users and stakeholders implicated in this effort, in order to reconcile some of the normative trade-offs identified in this paper.

Purposes and uses of growth data

The optimal cutoff point for a nutritional indicator depends, among other things, upon the purpose for which it is intended. This has been elaborated mathematically and conceptually elsewhere [1–4] and has stimulated some related empirical work [5–9]. A conceptual summary is included here, using mainly examples from young child nutrition where the concepts have been elaborated in earlier work. It should be noted that all of the principles illustrated in this chapter for undernutrition at the low end of anthropometric distributions apply to the high end, and vice versa.

Figure 1 illustrates seven categories of uses. Four of these (A–D) relate to targeting at the individual or population level.

» Indicators of past harm, such as stunting or underweight, are used to target programs to geographic or socioeconomic groups (at the population level) or to target services to individual children or households (at the individual level).
» Indicators of present deficits, such as dietary intake, are used to identify nutrient deficiencies or excess (at the population level) and sometimes as eligibility criteria for services (at the individual level, such as in the Women, Infants and Children Supplemental Food Program [WIC]).
» Predictors of future risk are used to target programs (e.g., weight-for-height in emergency situations or overweight in relation to chronic disease) and as eligibility criteria for individuals (e.g., low maternal weight or body-mass index [BMI] to target supplementary feeding to prevent low birthweight).
» Finally, in some settings anthropometric indicators have been used to derive socioeconomic indicators of deprivation, to assist targeting of programs by geographic or socioeconomic criteria (at the population level). As detailed elsewhere [1–3], the optimal cutoff point varies across these four targeting indicators.

As distinct from these four targeting indicators, all of which are related to past, present, or future risk of *harm or inequity,* **figure 1** also distinguishes indicators or predictors of *benefit.* That is, in most intervention situations there are some individuals (and, by extension, population groups) who may benefit from the intervention in question and some who may not (or may not benefit to the same extent), and certain indicators may serve as predictors of who is likely to benefit. Examples include age as a predictor of which children might benefit from supplementary feeding, plasma ferritin as a predictor of benefit from iron supplementation at the individual level, and the prevalence of malaria or hookworm at the population level as a predictor of population-level benefit from iron supplementation. Such predictors of benefit usually operate as effect modifiers or as factors that are correlated with effect modifiers.

Finally, the use of anthropometric indicators for *evaluation or monitoring* is a distinct purpose, whose primary requirement is that the indicator (and the chosen cutoff point) must be capable of *reflecting or responding* to the intervention, the program, or the changing conditions in question. For instance, responsive indicators might be the prevalence of very low energy intakes (in response to a supplementary feeding program targeted at the ultrapoor) or the incidence of intrauterine growth retardation (IUGR) in response to a targeted maternal health and nutrition program). A nonresponsive indicator might be stunting in relation to supplementary feeding. Responsiveness depends on certain characteristics of the indicator (i.e., its biological relationship to the intervention) as well as the

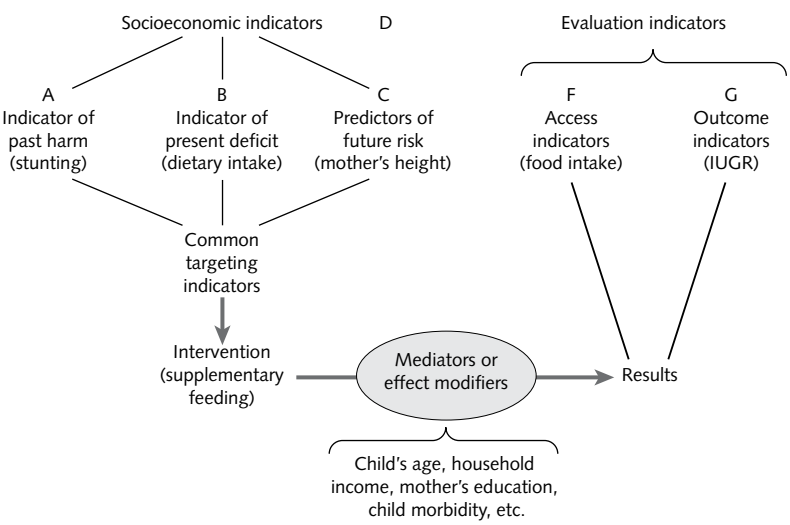

FIG. 1. Seven categories of uses for nutritional indicators. IUGR, intrauterine growth retardation

specific cutoff points used in the analysis.

This brief overview of distinct uses of anthropometric indicators highlights that there is a striking variety of uses of such indicators; the technically best indicator and the optimal cutoff point are likely to vary markedly across these uses; and apart from technical considerations, the choice of indicator and cutoff points also is, or should be, guided by a variety of resource, logistical, and administrative considerations.

Choosing cutoff points: A technical framework

Cutoff points can be defined on the basis of statistical, risk, and prescriptive criteria. This section compares these and shows that, although they do have some clear theoretical and philosophical differences, all three methods suffer from some of the same problems from a pragmatic perspective. This section uses cross-sectional references to illustrate some basic principles; longitudinal or mixed longitudinal references are affected by these same principles.

Figure 2A illustrates the statistical method, and **figure 2B** illustrates the underlying reality in that method. In developing a statistical reference, such as the National Center for Health Statistics (NCHS) reference, a nationally representative distribution is constructed from a healthy, well-nourished population, using only mild exclusion criteria to exclude overtly unhealthy or poorly nourished individuals. Statistical criteria are then used to define cutoff points, such as centiles or z-scores, as shown in **figure 2A**.

The weakness of the statistical method lies in the use of only mild exclusion criteria, which often is necessitated by lack of definitive information that might permit further exclusions. **Figure 2B** shows that the observed distribution actually is composed of two subdistributions: a healthy one (the Sp, or specificity, distribution) and an unhealthy one (the Se, or sensitivity, distribution). In the example shown, this results in a much larger variance (and, thus, more extreme cutoff points) than would have been the case had a truly "healthy, well-nourished population" been used as the reference. This can be seen by comparing the lower half of the observed mixed distribution with the lower half of the Sp distribution. The magnitude of the bias introduced into the cutoff points is in some proportion to the relative size (i.e., prevalence) of the unhealthy subpopulation (Se) versus the healthy subpopulation (Sp) and the distance between the Se mean and the Sp mean. Of particular interest is that the size of the error in estimating prevalence or classifying individuals as healthy versus unhealthy varies across the three cutoff points and in most cases the optimal cutoff point differs for these two purposes. Note that identical principles apply to overweight, in which case the distribution in **figure 2A** would have an extended tail at the upper end.

In developing a risk-based reference, epidemiologic studies are used to estimate the risk of an adverse outcome (or set of outcomes) at various levels of the anthropometric indicator. **Figure 3A** depicts a common risk curve from such studies, though a variety of functional forms are possible. **Figure 3B** shows that the reality underlying this risk curve is similar to that seen in the statistical criterion method. Specifically, the observed risk curve actually is a function of the relative size of the unhealthy subpopulation (Se) versus the healthy subpopulation (Sp) and the distance between the Se and Sp means. However, in this case it also is a function of the shape of the "true" underlying dose–response curve. As with the statistical method, the size of the error in estimating prevalence or classifying individuals varies across the three cutoff points shown, and the optimal cutoff typically differs for these two purposes.

Although the risk-based method has the attraction of providing a seemingly more objective and functional basis for defining a cutoff point, it also suffers from

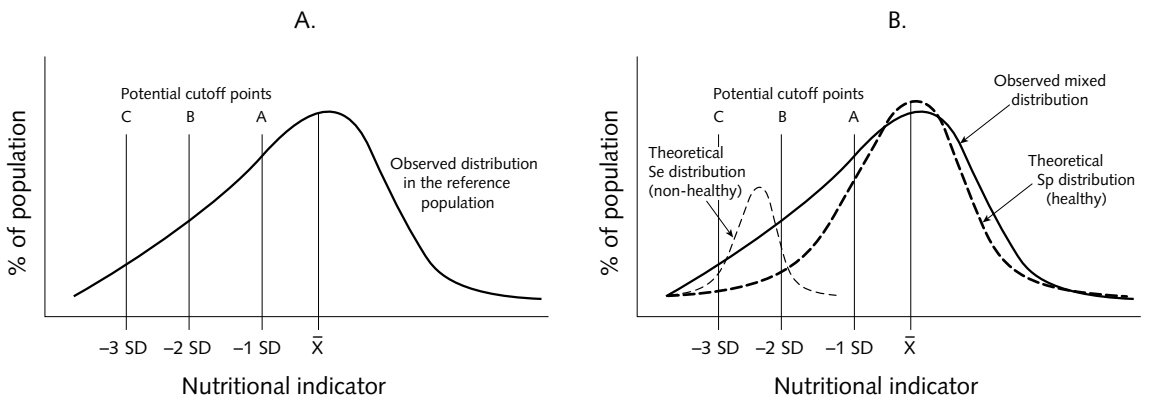

FIG. 2. A. Statistical method for defining cutoff points. B. Underlying reality in the statistical criterion method

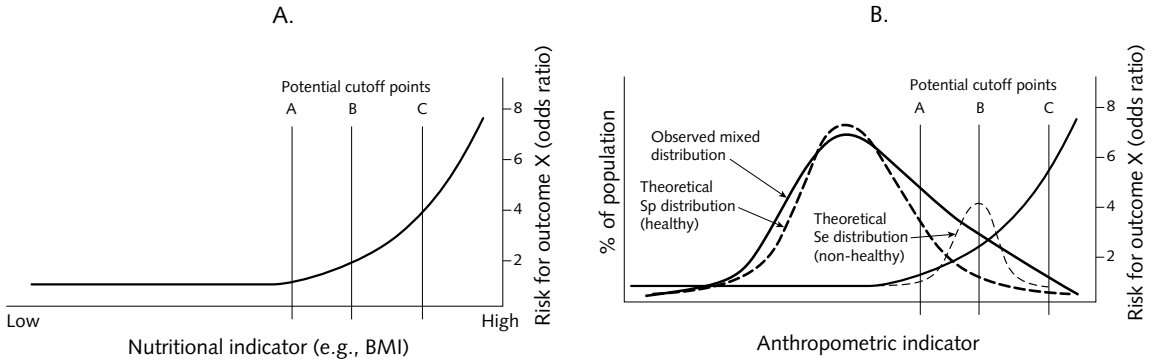

FIG. 3. A. Risk-based framework for defining cutoff points. BMI, body-mass index. B. Underlying reality in the statistical method at high end

several weaknesses, including the need to decide which outcome(s) should be included in the risk curve(s) and, if more than one outcome is relevant, how they should be aggregated or weighted; the need for empirical (often prospective) studies to estimate the functional form; a variety of misclassification errors and design problems that can bias the estimate of the functional form; and the possibility that the functional form as estimated in one population at one time may not by generalizable to other populations and times. A related problem is that reference values tend to be used for relatively long periods of time (e.g. a decade or more) because of the cost and disruption associated with changing them, but scientific understanding of risks and functional forms is continually evolving. Thus, a given set of risk-based reference values may become obsolete even before they have become widely adopted.

The prescriptive method seeks to define how children *should grow* rather than simply describe how they *do* grow. Conceptually, the construction of such a growth standard might begin with nationally representative samples from one or more suitable countries, which are modified by applying a variety of inclusion and exclusion criteria. These criteria may include behaviors and environmental conditions known or suspected to constrain growth (or cause excessive growth). They might also include some risk-related criteria reflecting the probability that a child of a given size may develop an adverse health or functional outcome some time in the future. The important point for the present purposes is that the development and performance of these criteria carry the same assumptions, information requirements, and weaknesses as those described for the statistical and risk-based methods. Conceptually, the process begins with an observed mixed distribution such as that shown in **figures 2A** and **3A** and then seeks to exclude as many of the "unhealthy" as possible while minimizing the accidental exclusion of the "healthy." Errors and inefficiencies in this process will adversely affect the mean, variance, and/or skewness of the distribution and, thus, under- or overestimate the location of the cutoff points.

Figure 4 illustrates that there are trade-offs in selecting and applying these criteria. For instance, if the sensitivity for predicting excessive growth as a function of five exclusion criteria is 50% and the specificity (for predicting normal growth) is 70%, then the "prescriptive distribution" will still be contaminated by 50% of those with excessive growth and will be missing 30% of those with normal growth. This will adversely affect the estimated variance and skewness of the distribution and thus the estimated location of the cutoff points. As the strictness of the exclusion criteria is increased, to maximize sensitivity, there is a corresponding decrease in specificity, and the remaining samples become less and less representative of the national population in terms of demographics, socioeconomics, ethnicity, and other factors correlated with these. These trade-offs were faced in constructing the Multicentre Growth Reference and led the committee to relax certain exclusion criteria in order to avoid basing the reference on a highly selective and potentially biased sample [10].

An important implication is that the validity of a prescriptive growth reference may *appear* to be greater when more strict inclusion criteria are applied, but this

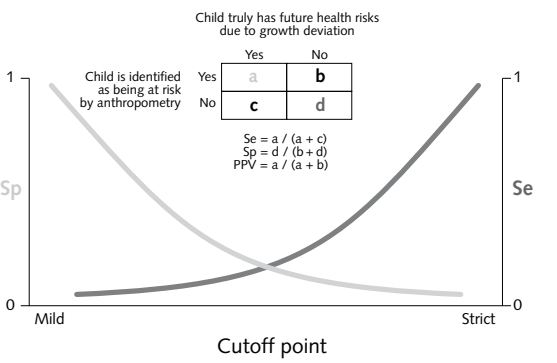

FIG. 4. Trade-offs in setting cutoff points for individual screening. Se, sensitivity; Sp, specificity; PPV, positive predictive value

comes at a cost. Ultimately, the validity depends on the strength of the statistical association and predictive accuracy between the exclusion criteria and growth (i.e., the perfect screening criterion would be perfectly correlated with the adverse outcomes) and the external validity of the studies or assumptions used in identifying these criteria. If the strictness, predictive accuracy, and external validity are low, the so-called prescriptive reference may simply be a slightly modified version of a statistical reference distribution, thereby weakening its claim as a normative standard.

In summary, this section reveals that the statistical, risk-based, and prescriptive methods all suffer from our relatively limited ability to distinguish the healthy and unhealthy subpopulations that underlie observed anthropometric distributions. The degree of distortion in the location of cutoff points variously depends upon the size of the healthy versus the unhealthy subpopulations, the difference in the means (and variances) of their respective distributions, inaccuracies in estimating the functional form of the risk curve, and the way in which trade-offs between sensitivity and specificity are resolved. Moreover, as shown elsewhere [1], these distortions have differential effects on applications at the individual level (such as screening) versus the population level (such as prevalence estimation).

Finally, it is relevant to note that some applications of anthropometry may be better served by comparing means and moments of distributions (among populations, over time, before and after interventions, etc.) rather than dichotomous variables or prevalence estimates [11]. When this is possible, it may avoid many of the complications associated with selecting cutoff points. Such a decision should be approached with care, however, because even the interpretation of means and moments may be complicated by population-level variation in non-nutritional factors such as maturation, environmental exposures, ethnicity, etc., and because in some situations changes in anthropometric characteristics may only be expressed in one of the tails of the distribution and may not be detected (as well) through comparison of means or moments alone. This phenomenon has been rarely observed in the case of indicators of chronic undernutrition among young children [11], but it may be relevant for indicators of overweight and adiposity among schoolchildren and adolescents.

High-end versus low-end cutoff points

The use of growth references in developing countries has emphasized a concern for detecting and avoiding constrained growth in children under 5 years of age, as reflected in height, weight, weight-for-height, and occasionally upper-arm circumference. The normative assumption has been that all children should be supported in reaching their genetic potential in size (notably height), and this has been operationalized as maintaining measurements within 2 SD above or below the reference median. There has been much less concern for "excessive growth" in these settings, as reflected in height, weight, BMI, adiposity, body fat distribution, or early maturation, but these are now concerns in developed and developing countries alike. Although all of the principles described above apply equally to high-end and low-end cutoff points, there are some additional considerations that have differential effects at the two ends of the distribution. These are discussed below in terms of technical and social considerations.

Technical considerations

As noted, all three methods (statistical, risk-based, and prescriptive) must confront the need to distinguish healthy from nonhealthy subpopulations, though they do so in different ways. This has different implications for the upper and lower segments of the distribution. Specifically:

- The health or functional consequences of low-end deviations are different in character (e.g., greater risk of infectious disease morbidity and mortality, impaired cognitive development) than those of high-end deviations (greater risk of diabetes, hypertension, heart disease, and other later-onset chronic diseases);
- These outcomes generally are manifested and can be detected earlier for low-end deviations (e.g., months or years versus decades);
- As a result of the above, there is greater statistical uncertainty in estimating the dose–response relationship at the high end than at the low end;
- It is more difficult to define normatively desirable behaviors at the high end (and, thus, inclusion and exclusion criteria) in the context of the prescriptive method; this is in contrast to desirable behaviors at the low end, which may include exclusive breastfeeding, appropriate complementary feeding, and absence of illness. In the case of high-end deviations in BMI (and the dietary and physical activity behaviors that might be considered "desirable" in creating a prescriptive reference), there are strong personal, social, cultural, and political values at stake, such that the creation of a prescriptive reference based on a narrow set of values in the face of weak or uncertain scientific criteria may be strongly contested;
- The "acceptable" degree of lability or reversibility in measurement may differ at the high and the low ends, because of biological limits to catch-up growth in impoverished settings, greater plasticity in soft tissue traits, and uncertainty or unknowns concerning the health consequences of various growth patterns (e.g., weight cycling) at the high end versus the low end.

The implications of these considerations can be seen in **figure 5** and **table 1**. In this hypothetical

example based on risks as they might be observed in a population from which a reference or standard is to be constructed, the midpoint dose–response curve is steeper for an outcome at the low end of the reference distribution than for an outcome at the high end, but the standard error of the estimate is greater at the high end. Thus, a given statistical cutoff point (such as –2 SD versus +2 SD) is associated with greater health risks at the low end (odds ratio, 2.5) than at the high end (odds ratio, 1.6) when the midpoint risk estimate is used; however, if the risks associated with these same cutoff points were to be based on parameter estimates at +1 SE (to increase sensitivity in a risk-based reference or to enforce strict inclusion criteria in a prescriptive reference), in this hypothetical case these cutoff points would have virtually the same epidemiologic meaning (odds ratio, 2.8 versus 2.5, respectively).

The relevant insight from the above is that the distinct health or functional implications of high-end versus low-end deviations in anthropometry, and the asymmetries in uncertainties, make it theoretically as well as practically very difficult to define cutoff points that have a common epidemiologic meaning with respect to Se, Sp, and prevalence, or a comparable meaning in terms of health or function. These complications affect all three methods of constructing a reference, but they are especially pronounced for the risk-based and prescriptive methods, because the claims and aspirations of these methods may exceed the underlying evidence base and current levels of predictive accuracy in defining "healthy" or "desired"

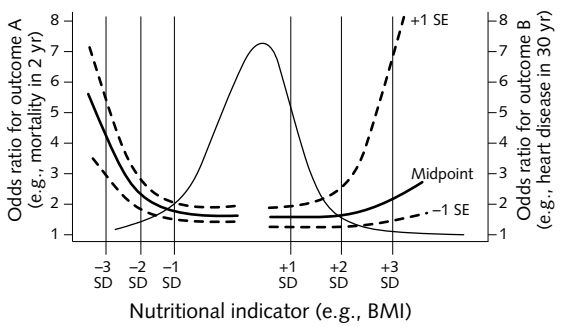

FIG. 5. Asymmetries in risks and uncertainties at low and high end. BMI, body-mass index

TABLE 1. Parameter estimates of hypothetical health risks at the low end and the high end of an indicator[a]

Parameter estimate	Low-end cutoff points			High-end cutoff points		
	–1	–2	–3	+1	+2	+3
–1 SE	1.5	1.8	2.5	1.0	1.0	1.2
Midpoint	1.8	2.5	3.8	1.5	1.6	2.0
+1 SE	2.0	2.8	5.0	1.8	2.5	6.0

a. Values are odds ratios, taken from **figure 5.**

size or growth. In addition, the use of high end versus the low end of the distributions carries important social implications, as described next.

Social considerations

The need to address both low-end and high-end deviations in anthropometric characteristics introduces considerable complexity from a social perspective. This is because each of the technical issues described above has underlying social implications; low-end deviations typically are perceived to carry different social connotations than high-end deviations; and the politics of problem definition and policy development can differ for the low-end and high-end deviations and can place added scrutiny on the technical and normative foundations for the reference or standard.

The technical issues described in the previous section, notably the asymmetries in risks and scientific uncertainties, suggest a need to consider the differential social implications of errors at the two ends of the distribution. Moreover, there is a need to do so in light of each of the uses of anthropometry at the individual and population level. To illustrate on the basis of **figure 5** and **table 1**, one might use a common statistical criterion for choosing the cutoff point (e.g., –2 SD for the low end and +2 SD for the high end). This may appear even-handed from a technical or statistical perspective, but the performance of these cutoff points for use in clinical screening (viz. Se and Sp) actually will be quite different, because the odds ratio is 2.5 in the former case and 1.6 in the latter case (at the midpoint). Similarly, the appearance of even-handedness in prevalence estimation obscures the underlying reality that a higher proportion of the "undernourished" (as defined by the –2 SD cutoff point) than of the "overnourished" (as defined by the +2 SD cutoff) are at risk for impairment of health or function. In effect, the use of a common statistical criterion for the cutoff at the two ends of the distribution amounts to a systematic overestimation of the prevalence at the high end (relative to true functional outcomes).

One option for avoiding such difficulties might be to choose the cutoff points that will yield an odds ratio at the low end identical to that at the high end, so that they will have a common health or functional interpretation when used in screening or prevalence estimation.* However, this introduces additional complications. In the example, a statistically more extreme cutoff point would be required at the high end than at the low end (i.e., roughly +4 SD versus –2 SD), and the standard error of the estimate is far greater in that region. This could entail even greater errors in screen-

* The technically more correct procedure would be to choose a cutoff where the sum of Se and Sp is maximal [2], but this entails the same complications as a common odds ratio.

ing and prevalence estimation and would be very difficult to explain and justify to users.

These considerations raises several social/normative questions:

» What are the private and social benefits, costs, and consequences of each of these outcomes? How much error can be tolerated in this process, viz. false low-end positives, false high-end positives, false low-end negatives, and false high-end negatives? What are the private and social costs and consequences of these errors?

» What are the social/normative, health, and political implications of overestimating or underestimating undernutrition in relation to overnutrition (at the individual or the population level)?

A related technical question with important social implications relates to the possibility that catch-up growth and size as an adult may have a different biological and health meaning, depending on the nature of early nutritional experiences [12]. Although this possibility requires further study, it could call into question the appropriateness of a single international growth standard, because the answer to how children (as individuals and populations) *should* grow would depend upon when and how they *did* grow during gestation and infancy.

In addition to the technical factors that give rise to such questions, the social perception of low-end versus high-end deviations in anthropometric characteristics differs in significant ways, related to causality (or blame) and policy solutions (or responsibility). To what extent is the problem (i.e., undernutrition or overnutrition) a result of behavioral versus environmental factors? Regardless of the answer to this question, to what extent should the responsibility for change lie with the individual versus the community or society. These themes form the social, cultural, and political background against which the "error trade-offs" posed above are implicitly evaluated. It may be necessary for the the developers of anthropometric standards to articulate an explicit position on these issues so that it can reconcile the error trade-offs when recommending cutoff points.

Finally, events in recent years suggest the possibility that interest-group politics may lead to greater scrutiny of the technical and normative basis for defining overweight and obesity than has been the case in recent years with undernutrition. These events include the controversy attending the release of the WHO report on "Diet, Nutrition and the Prevention of Chronic Diseases" [13]; the CDC paper revising estimates of mortality associated with overweight and obesity [14]; the significant media, legislative, and corporate attention devoted to the subject; and the increasingly politicized debates concerning evidence-based decision-making and sound science in policy development [15]. Should interest-group politics become focused on the definition of overweight and obesity, they are likely to focus on the justification for a prescriptive reference rather than a comparative reference, the countries chosen to define that norm, and the strength of evidence underlying decisions regarding inclusions, exclusions, and cutoff points. It also is possible that the debate, once begun, could spill over into the low end of the distribution. These possibilities further suggest a need to pay explicit attention to the technical and normative basis for reconciling the error trade-offs noted above and to revisit the advisability of advancing a normative standard at both ends of the distribution.

Illustration of principles based on body-mass index (BMI)

Although the above principles have been advanced entirely at a conceptual level for ease of presentation, it is instructive to examine them with a specific example in mind. This section illustrates some of the principles using the high-end cutoff points for BMI of children and adolescents as an example.

Historically, the NCHS references used the 85th and 95th age- and sex-specific percentiles of BMI as cutoff points for overweight and obesity during childhood and adolescence [16]. As such, these represent statistically based cutoff points. However, with the increasing concern regarding the long-term health consequences of obesity, and the documented rightward shift in the upper tails of the BMI distribution among US children and adolescents, the use of these percentiles based on US children, even as reference values, became less tenable. The continued use of these cutoff points effectively would have "normalized" these higher BMIs and provided a biased yardstick for estimating overweight and obesity in the United States and in other populations.

As an alternative, the International Obesity Task Force (IOTF) developed a set of reference cutoff points that, statistically, correspond to a BMI of 25 and 30 as observed among adults in a pooled international sample [17]. Those cutoff points, in turn, have been shown in earlier studies to be associated with elevated health risks among adults [18]. In choosing this strategy, the IOTF recognized that a given cutoff point in adults may not carry the same health implications among children and adolescents, but they decided that the errors associated with this strategy were preferable to those associated with continued use of the 85th and 95th US centiles. Thus, the IOTF cutoff points are a mix of statistical, risk-based, and pragmatic considerations. An important question is the extent to which child and adolescent BMIs above these or any other cutoff points are associated with future health risks.

Epidemiologic studies among children and adolescents have used a variety of cutoff points to examine the future health risks associated with elevated BMI,

making comparisons difficult. However, the literature on longitudinal studies is consistent at a general level in showing statistically significant associations between elevated BMI and future health risks, including elevated risks of all-causes and cardiovascular mortality [19–21], arterial (intimal) medial thickening [22–24], coronary artery calcification [25], atherosclerosis [26], metabolic syndrome [27–29], fasting glucose and insulin [24, 29], and blood pressure and blood lipids [28, 29]. These findings support the view that the increasing prevalence of elevated BMI among children and adolescents is likely to increase the *population-level* burden of chronic disease in the future.

For the purpose of estimating the *population-level* burden of chronic disease in the future, as summarized in the population-attributable risk (PAR), for instance, the most appropriate cutoff points would be located where the risk curve first increases above baseline and at each inflection point in the curve. This might be at points A, B, and C in **figure 3A**, for instance. Alternatively, PAR could be calculated (and calculated more precisely) without the use of cutoff points by simply combining the risk equation with the distribution of BMI values in a population [30]. Such calculations generally lead to the conclusion that overweight and obesity have a large impact on the burden of chronic disease at the population level [31–34].

A basic principle outlined in this article is that a given set of cutoff points may perform well for population-level uses but may or may not perform well for individual-level uses. In the former case, as in calculating PAR, it is often sufficient to show that there is a consistent and statistically significant association between BMI and future health risks, as summarized in the odds ratio, for instance. In the latter case, it is necessary to use different performance measures, such as sensitivity, specificity, and positive predictive value, to assess the predictive accuracy at the individual level.

Few long-term longitudinal studies are available that examine the future health risks of elevated BMI among children and adolescents, and fewer still have published findings in the form necessary to permit their predictive accuracy to be evaluated. **Table 2** presents the findings from three studies that do permit such measures to be derived from the published results. Several findings are noteworthy:

» There is consistency across studies, cutoff points, and outcome variables in showing significant odds ratios between elevated BMIs and future health risks;
» There is considerable variation in sensitivity and positive predictive value across different outcomes and cutoff points, and generally high specificity in all cases. Odds ratios (and regression coefficients in other analyses) are not necessarily closely related to these measures of predictive accuracy;
» There is wide variation in the ratio of false positives to true positives (b/a in **fig. 4**), but in most cases use of these cutoff points would yield three to five false positives for each true positive identified. The FP/TP ratio is not necessarily related to the strength of the odds ratio or Se; in fact, some of the highest FP/TP ratios are found in conjunction with the highest odds ratios.*
» Although the metabolic syndrome is closely associated with BMI in terms of odds ratios and Se, the use of these BMI cutoff points still yields very high FP/TP ratios in the one Bogalusa study [28] that did not include waist circumference in the definition of that syndrome.
» The one mortality study for which these statistics could be calculated, relating BMI in childhood or adolescence to adult mortality, suggests similar findings to those for the risk factors, based on a relatively low BMI cutoff point (> 75th percentile). Although the full range of statistics cannot be calculated from the paper, **table 3** shows that the strength of the association (odds ratio) between *adult* BMI and future adult mortality (from a meta-analysis of 26 studies) is not statistically significant for BMIs in the overweight range (25 to 29) and is significant, but not impressive in absolute terms, in the obesity range (> 30). This is so for the median estimate of the odds ratio as well as the upper limit of the 95% confidence interval in the meta-analysis [36].

This example confirms that the performance of cutoff points can vary in important ways between population-level uses (such as estimating PAR) and individual-level uses (such as screening). The literature tends to emphasize the reporting of odds ratios and regression coefficients, by way of demonstrating statistically significant associations (and possibly causality), with much less attention to the performance measures bearing on predictive accuracy. The available data suggest odds ratios are not particularly informative concerning the predictive accuracy of the cutoff point. Se and Sp are more informative, and are widely recognized as such in clinical epidemiology. However, the present example highlights another performance measure, the FP/TP ratio, which is generally overlooked in the extant literature and speaks directly to a potentially very important value trade-off in choosing cutoff points for overweight and obesity.** This ratio is directly affected by the placement of the cutoff point, and the relatively low BMI cutoff (25 kg/m^2) used in

* The false positive *ratio* as calculated here is different from the false positive rate more commonly used in clinical epidemiology. The latter is defined as b/(b+d) or (1–Sp). There are inconsistencies in how these are used in the literature [35].

** The FP/TP ratio is closely related to the positive predictive value (PPV), in that FP = 1–PPV. However, the PPV implicitly emphasizes a desire to detect as many true positives as possible, whereas the FP/TP ratio forces an explicit ethical comparison between that goal (related to beneficence) and two competing goals (nonmaleficence and the desire to focus intervention resources on those truly in need or at risk).

TABLE 2. Predictive accuracy of risk factors and mortality in early adulthood based on BMI in childhood and adolescence

BMI range	Beginning age (yr)	Follow-up age (yr)	Outcome variable	OR	Se	Sp	PPV	FP/TP ratio	Study
25–29	4–15	27–39	Hypertension	2.65	35.9	85.2	5.4	17.5	Bogalusa [28]
			High LDL cholesterol	2.18	25.9	85.9	20.0	4.0	
			Low HDL cholesterol	2.13	24.6	87.1	34.4	1.9	
			High triglycerides	1.99	26.5	85.9	20.0	4.0	
			Metabolic syndrome[a]	3.69	38.1	87.0	23.0	3.3	
> 30	4–15	27–39	Hypertension	4.65	17.9	96.1	9.7	9.3	
			High LDL cholesterol	1.73	6.3	96.0	14.5	5.9	
			Low HDL cholesterol	2.37	7.1	96.5	31.6	2.2	
			High triglycerides	2.13	7.2	96.0	16.0	5.2	
			Metabolic syndrome	4.87	12.5	96.6	28.0	2.6	
> 95p	5–17	24–36	High systolic BP	4.5	34	—	13	6.7	Bogalusa [29]
			High diastolic BP	2.4	23	—	9	10.1	
			High LDL cholesterol	3.0	28	—	18	4.6	
			Low HDL cholesterol	3.4	25	—	17	4.9	
			High triglycerides	7.1	47	—	24	3.2	
			High insulin	12.6	62	—	21	3.8	
			1+ RF	2.8	25	89	49	1.0	
			2+ RF	7.0	48	88	27	2.7	
			3+ RF	17.9	74	86	10	9.0	
	5–10	24–29	1+ RF	5.4	29.4	92.9	60.6	0.65	
			2+ RF	8.8	50	90	26	2.8	
			3+ RF	29.4	80	88	12	7.3	
	11–17	29–36	1+ RF	1.9	22	87	42	1.38	
			2+ RF	6.2	48	87	27	2.7	
			3+ RF	13.7	70	85	10	9.2	
> 75p	2–14	59–72	All-cause mortality	1.5	13.5	86.6	15.7	5.4	Boyd-Orr [21]
			CHD mortality	2.0	13.1	86.6	.046	20.7	

BMI, body-mass index; OR, odds ratio; Se, sensitivity; Sp, specificity; PPV, positive predictive value; FP, false positive; TP, true positive; LDL, low-density lipoprotein; HDL, high-density lipoprotein; BP, blood pressure; RF, risk factor; CHD, coronary heart disease

a. Metabolic syndrome is defined by abnormal levels of three or more of the following seven risk factors: systolic blood pressure, diastolic blood pressure, triglycerides, LDL cholesterol, HDL cholesterol, fasting glucose level, and waist circumference. Shading added for emphasis.

TABLE 3. Predictive accuracy of adult mortality based on BMI in adulthood: meta-analysis of 26 studies [36]

BMI range	Beginning age	Follow-up age	Outcome variable	OR[a]
25–29	Adult males	Adult males	All-cause mortality	0.97 (1.01)
			CHD mortality	1.16 (1.16)
	Adult females	Adult females	All-cause mortality	0.97 (0.99)
			CHD mortality	1.10 (1.20)
> 30	Adult males	Adult males	All-cause mortality	1.20 (1.20)
			CHD mortality	1.51 (1.67)
	Adult females	Adult females	All-cause mortality	1.27 (1.29)
			CHD mortality	1.62 (1.81)

BMI, body-mass index; OR, odds ratio; CHD, coronary heart disease

a. Values in the odds ratio column represent the summary relative risks (with the reference BMI group being 18.5 to 24.9). The upper limit of the 95% confidence interval based on all 26 studies is shown in parentheses.

estimating PAR, for instance, may perform poorly in terms of the FP/TP ratio. In light of the concern and growing evidence regarding the social and psychological dimensions of child and adolescent weight status in some contexts [37–42], this ratio should be included with other measures of predictive accuracy when alternative cutoff points are evaluated.

A common justification for using low cutoff points, such as a BMI of 25, despite the relatively weak association with health outcomes, is that modest elevations in risk in the BMI range of 25 to 30 can lead to substantial morbidity, mortality, and associated outcomes at the population level. Although this often is a valid claim, the above results further show that the use of the same cutoff points for screening purposes may result in poor predictive accuracy, with associated poor identification of those at risk (due to low Se), leakage of scarce intervention resources (due to low PPV), and potential unintended consequences (due to high FP/TP). When the population-level uses of the data are used to implement *exclusively* population-level public health actions (such as fortification, improving water and sanitation or the built environment, imposing taxes, etc.) there may be no conflict in the use of such cutoff points to report prevalence, PAR, or other statistics. However, these instances are rare. In most cases, so-called public health strategies do include a component of population-wide *individual* actions, such as growth-monitoring programs, school or worksite screening programs, public awareness and education campaigns, and more generalized effects on individuals transmitted through the media and informal cultural channels. In such cases, it is appropriate to consider the potential unintended consequences associated with false positives. This suggests that there can be a trade-off, or conflict, involved in using mild cutoff points for BMI (and other indicators) at the population level when their predictive accuracy is low at the individual level.

An integrated framework and key issues for discussion

This paper, and others in this series, have identified several categories of considerations relevant to the choice of cutoff points, as depicted in **table 4**. This table underscores the need to distinguish children from adolescents and, in each case, the low end from the high end of the distribution. In addition, it identifies a set of complicating factors to be considered in each case. These factors include some that often are treated as "intrinsic factors" (shaded in the table) but that actually are nonseparable from environmental factors; a large set grouped under behavioral and environmental variation; the existence of variability and uncertainty in estimating the causes and consequences (and slopes) of anthropometric variation; and the special challenges of constructing cross-sectional versus longitudinal references. Finally, the table underscores that the most efficient indicator and cutoff point typically varies across the many purposes to which anthropometric data are applied.

If taken literally, **table 4** implies the need to develop an entire family of cutoff points, which clearly is neither possible with the existing knowledge and evidence nor practical from a user's perspective. However, the table provides a comprehensive framework to guide decisions about key issues requiring attention. The following questions and suggestions are offered for further discussion:

1. In contrast to the preschool period, when "reaching an individual's genetic potential" has served as a touchstone for developing a reference, a different theoretical or philosophical foundation may be needed during childhood and adolescence. This is because of concerns both about the causes of early maturation (such as dietary and environmental exposures or overweight itself) and about its consequences (such as early and prolonged exposure to endogenous estrogen, which is linked to breast cancer). The new growth reference may need to question the notion that bigger is better, which is obvious in relation to weight status but also may extend to the case of height, if elevated height-for-age is merely a reflection of early maturation resulting from adverse dietary or environmental exposures and/or overweight.
2. In light of the technical and social challenges associated with specifying prescriptive behaviors and estimating long-term health effects at the high end of anthropometric indicators, consideration could be given to developing a prescriptive reference at the low end and a comparative reference at the high end. Consideration could also be given to using samples from different countries for the low-end reference versus the high-end reference.
3. In light of the inefficiencies and error trade-offs associated with the use of single cutoff points at both ends of an anthropometric distribution, consideration could be given to the development of two cutoff points at each end. The more extreme cutoff point could carry stronger normative claims and health implications. The less extreme cutoff point would signal the need for further assessment and caution (at the individual level) and would provide a basis for estimating the size of the at-risk group at the population level. The gain in efficiency and predictive power from such an approach, at the individual and population levels, might be estimated by using available studies and reanalysis of existing datasets. The acceptability, comprehensibility, and credibility of such a system could be tested through consultation with end users and stakeholders with diverse viewpoints on these issues.

TABLE 4. Considerations in selecting anthropometric cutoff points[a, b]

Age group (yr)	Complicating factor	Screening		Prevalence			Responsiveness	
		On risk	On benefit	Single population	Compare populations	Temporal changes	Individual	Population
Low end of distribution								
5–9	Age							
	Height/allometry							
	Maturity (mid-spurt)							
	Ethnicity							
	Genetics							
	Behavioral and environmental variation							
	Variability and uncertainty of causes, consequences, and slopes							
	Cross-sectional vs. longitudinal study							
10–18	Age							
	Height/allometry							
	Maturity (mid-spurt)							
	Ethnicity							
	Genetics							
	Behavioral and environmental variation							
	Variability and uncertainty of causes, consequences, and slopes							
	Cross-sectional vs. longitudinal study							
High end of distribution								
5–9	Age							
	Height/allometry							
	Maturity (mid-spurt)							
	Ethnicity							
	Genetics							
	Behavioral and environmental variation							
	Variability and uncertainty of causes, consequences, and slopes							
	Cross-sectional vs. longitudinal study							
10–18	Age							
	Height/allometry							
	Maturity (mid-spurt)							
	Ethnicity							
	Genetics							
	Behavioral and environmental variation							
	Variability and uncertainty of causes, consequences, and slopes							
	Cross-sectional vs. longitudinal study							

a. The table illustrates conceptually the full range of theoretical considerations that may affect choice of cutoff point; each cell could contain a different cutoff point.
b. Shading indicates intrinsic factors.

4. Consideration could be given to developing a system for risk-based exclusion criteria. Specifically, one might exclude a proportion of children from the reference sample or samples on the basis of the probability that children of their size will develop adverse outcomes in the future. The latter would be estimated from a meta-analysis of existing studies, with appropriate weighting and aggregation across multiple outcomes. This also may provide the epidemiologic evidence needed to determine the optimal cutoff points described in paragraph 3 above.

5. In light of the trade-offs among Se, Sp, PPV, and FP/TP documented here, the normative (value-laden) implications of any reference (but especially a prescriptive one) should be explored through in-depth consultation with an appropriately diverse set of end users and stakeholders, rather than by public health specialists alone. Such a procedure is consistent with theoretical and philosophical considerations and also could prove vital in defending the normative status of such a standard.

References

1. Habicht JP. Some characteristics of indicators of nutritional status for use in screening and surveillance. Am J Clin Nutr 1980;33:531–5.
2. Habicht JP, Meyers LD, Brownie C. Indicators for identifying and counting the improperly nourished. Am J Clin Nutr 1982;35(5 suppl):1241–54.
3. Habicht JP, Pelletier DL. The importance of context in choosing nutritional indicators. J Nutr 1990;120(suppl 11):1519–24.
4. Brownie C, Habicht JP. Selecting a screening cut-off point or diagnostic criterion for comparing prevalences of disease. Biometrics 1984;40:675–84.
5. Ruel M, Rivera J, Habicht JP, Martorell R. Differential response to early nutrition supplementation: long-term effects on height at adolescence. Int J Epidemiol 1995;24:402–12.
6. Brownie C, Habicht JP, Cogill B. Comparing indicators of health or nutritional status. Am J Epidemiol 1986;124:1031–44.
7. Rasmussen K, Habicht JP. Malnutrition among women: indicators to estimate prevalence. Food Nutr Bull 1989;11:29–37.
8. Ross J. Patterns and determinants of maternal nutritional status during lactation in Malawi. Ithaca, NY, USA: Cornell University, Division of Nutritional Sciences, 1995. PhD Thesis
9. Du L. Fortifying Chinese soy sauce with iron: a study of the scientific and policy aspects of a food fortification program. Ithaca, NY, USA: Cornell University Division of Nutritional Sciences, 2005. PhD thesis.
10. de Onis M, Garza C, Victora CG, Onyango A, Frongillo EA, Martines J. The WHO Multicentre Growth Reference Study (MGRS): Planning, study design, and methodology. Food Nutr Bull 2004;25 (suppl 1):S15–26.
11. Yip R, Scanlon K. The burden of malnutrition: a population perspective. J Nutr 1994;124(10 suppl):2043S–6S.
12. Victora CG, Barros FC. Commentary: The catch-up dilemma—relevance of Leitch's "low-high" pig to child growth in developing countries. Int J Epidemiol 2001; 30:217–20.
13. Diet, nutrition and the prevention of chronic diseases. World Health Organ Tech Rep Ser 2003;916:1–149.
14. Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. JAMA 2005;293:1861–7.
15. Scientific integrity in policymaking. Cambridge, Mass, USA: Union of Concerned Scientists, 2004. Available at: http://www.ucsusa.org/scientific_integrity/. Accessed 6 September 2006.
16. Barlow SE, Dietz WH. Obesity evaluation and treatment: Expert Committee recommendations. The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services. Pediatrics 1998;102:E29.
17. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240–3.
18. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:i–xii, 1–253.
19. Dietz WH. Childhood weight affects adult morbidity and mortality. J Nutr 1998;128:411S–4S.
20. Must A, Jacques PF, Dallal GE, Bajema CJ, Dietz WH. Long-term morbidity and mortality of overweight adolescents. A follow-up of the Harvard Growth Study of 1922 to 1935. N Engl J Med 1992;327:1350–5.
21. Gunnell DJ, Frankel SJ, Nanchahal K, Peters TJ, Davey Smith G. Childhood obesity and adult cardiovascular mortality: a 57-y follow-up study based on the Boyd Orr cohort. Am J Clin Nutr 1998;67:1111–8.
22. Li S, Chen W, Srinivasan SR, Bond MG, Tang R, Urbina EM, Berenson GS. Childhood cardiovascular risk factors and carotid vascular changes in adulthood: the Bogalusa Heart Study. JAMA 2003;290:2271–6.
23. Raitakari OT, Juonala M, Kahonen M, Taittonen L, Laitinen T, Maki-Torkko N, Jarvisalo MJ, Uhari M, Jokinen E, Ronnemaa T, Akerblom HK, Viikari JS. Cardiovascular risk factors in childhood and carotid artery intima-media thickness in adulthood: the Cardiovascular Risk in Young Finns Study. JAMA 2003;290:2277–83.
24. Davis PH, Dawson JD, Riley WA, Lauer RM. Carotid intimal-medial thickness is related to cardiovascular risk factors measured from childhood through middle age: The Muscatine Study. Circulation 2001;104:2815–9.
25. Mahoney LT, Burns TL, Stanford W, Thompson BH, Witt JD, Rost CA, Lauer RM. Coronary risk factors measured in childhood and young adult life are associated with coronary artery calcification in young adults: the Muscatine Study. J Am Coll Cardiol 1996;27:277–84.
26. Berenson GS, Srinivasan SR, Bao W, Newman WP 3rd, Tracy RE, Wattigney WA. Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study. N Engl J Med 1998;338:1650–6.
27. Chen W, Srinivasan SR, Li S, Xu J, Berenson GS. Metabolic syndrome variables at low levels in childhood are beneficially associated with adulthood cardiovascular risk: the Bogalusa Heart Study. Diabetes Care 2005;28:126–31.
28. Janssen I, Katzmarzyk PT, Srinivasan SR, Chen W, Malina RM, Bouchard C, Berenson GS. Utility of childhood BMI in the prediction of adulthood disease: comparison of national and international references. Obes Res 2005;13:1106–15.

29. Freedman DS, Dietz WH, Srinivasan SR, Berenson GS. The relation of overweight to cardiovascular risk factors among children and adolescents: the Bogalusa Heart Study. Pediatrics 1999;103(6 pt 1):1175–82.
30. Pelletier D, Frongillo EA, Shroeder DG, Habicht JP. A methodology for estimating the contribution of malnutrition to child mortality in developing countries. J Nutr 1994;124:2106S–22S.
31. Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. JAMA 2005;293:1861–7.
32. Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB. Annual deaths attributable to obesity in the United States. JAMA 1999;282:1530–8.
33. Banegas JR, Lopez-Garcia E, Gutierrez-Fisac JL, Guallar-Castillon P, Rodriguez-Artalejo F. A simple estimate of mortality attributable to excess weight in the European Union. Eur J Clin Nutr 2003;57:201–8.
34. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States. JAMA 2004; 291:1238–45.
35. Suojanen J. False false positive rates. N Engl J Med 1999; 341:131.
36. McGee DL; Diverse Populations Collaboration. Body mass index and mortality: a meta-analysis based on person-level data from twenty-six observational studies. Ann Epidemiol 2005;15:87–97.
37. Fallon EM, Tanofsky-Kraff M, Norman AC, McDuffie JR, Taylor ED, Cohen ML, Young-Hyman D, Keil M, Kolotkin RL, Yanovski JA. Health-related quality of life in overweight and nonoverweight black and white adolescents. J Pediatr 2005;147:443–50.
38. Swallen KC, Reither EN, Haas SA, Meier AM. Overweight, obesity, and health-related quality of life among adolescents: the National Longitudinal Study of Adolescent Health. Pediatrics 2005;115:340–7.
39. Zeller MH, Saelens BE, Roehrig H, Kirk S, Daniels SR. Psychological adjustment of obese youth presenting for weight management treatment. Obes Res 2004;12: 1576–86.
40. Tershakovec A. Psychological considerations in pediatric weight management. Obes Res 2004;12:1537–8.
41. Williams J, Wake M, Hesketh K, Maher E, Waters E. Health-related quality of life of overweight and obese children. JAMA 2005;293:70–6.
42. Falkner NH, Neumark-Sztainer D, Story M, Jeffery RW, Beuhring T, Resnick MD. Social, educational, and psychological correlates of weight status in adolescents. Obes Res 2001;9:32–42.

The International Growth Standard for Preadolescent and Adolescent Children: Statistical considerations

T. J. Cole

Abstract

This article discusses statistical considerations for the design of a new study intended to provide an International Growth Standard for Preadolescent and Adolescent Children, including issues such as cross-sectional, longitudinal, and mixed designs; sample-size derivation for the number of populations and number of children per population; modeling of growth centiles of height, weight, and other measurements; and modeling of the adolescent growth spurt. The conclusions are that a mixed longitudinal design will provide information on both growth distance and velocity; samples of children from 5 to 10 sites should be suitable for an international standard (based on political rather than statistical arguments); the samples should be broadly uniform across age but oversampled during puberty, and should include data into adulthood. The LMS method is recommended for constructing measurement centiles, and parametric or semiparametric approaches are available to estimate the timing of the adolescent growth spurt in individuals. If the new standard is to be grafted onto the 2006 World Health Organization (WHO) reference, caution is needed at the join point of 5 years, where children from the new standard are likely to be appreciably more obese than those from the WHO reference, due to the rising trends in obesity and the time gap in data collection between the two surveys.

Key words: Growth charts, puberty, sample size, survey design

The author is affiliated with the Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College London, London.

Please direct queries to the author: T. J. Cole, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK; e-mail: tim.cole@ich.ucl.ac.uk.

Introduction

The World Health Organization (WHO) Multicentre Growth Reference Study (MGRS) has been progressing since 1999 [1], and the first charts resulting from the study were published in April 2006. The study has involved infants followed longitudinally from birth to 24 months and children sampled cross-sectionally between 18 and 71 months, from six sites across the world. Strict inclusion criteria have ensured that the infants, and to a lesser extent the children, have experienced unconstrained growth. The charts constitute an international growth standard covering the age range from birth to 5 years (see http://www.who.int/childgrowth/en/).

The international growth standard for preadolescent and adolescent children is now proposed to extend the WHO MGRS from 5 years through to puberty and beyond. For the purposes of this article, the philosophy of the international growth standard for preadolescent and adolescent children is assumed to match that of the MGRS, i.e., prescriptive subject selection to provide a standard for growth. However, several important aspects of the study design remain to be decided, including the upper end of the age range and whether to focus on cross-sectional (distance) or longitudinal (velocity) data.

Cross-sectional, longitudinal, and mixed designs

The first question to consider is whether to collect information cross-sectionally or longitudinally. A cross-sectional design involves measuring children on a single occasion, whereas a longitudinal design follows children over time and measures them repeatedly. A mixed longitudinal design combines features of both cross-sectional and longitudinal designs, measuring some children once and others more than once. A cross-sectional design estimates growth *distance* (i.e., size), whereas longitudinal or mixed longitudinal designs provide information on both growth distance and growth *velocity*.

Most current national growth references, such as the US Centers for Disease Control and Prevention (CDC) 2000 [2], British 1990 [3], and Dutch [4] references, are based on cross-sectional data. However, in the past, longitudinal growth studies, such as the French [5] and British [6, 7] studies, recruited children at birth and followed them through to maturity. In general, a longitudinal study needs to follow the same subjects throughout the age range of interest, which implies a study period of 10 years or more in the present context. However, depending on the time interval over which velocity is to be measured, a mixed design can reduce this period substantially. Velocity in childhood is typically measured over 1 year [6, 7], so that after an initial cross-sectional survey, the required longitudinal information can be obtained by remeasuring a fraction of the original sample 1 year later. A new cross-sectional sample can also be measured on the second occasion if required. This means that conducting two surveys one year apart permits the estimation of both distance and annual velocity throughout the age range.

More extended estimates of velocity inevitably require a longer period of follow-up for each individual. If, for example, the pubertal growth spurt is of interest, then each subject will need to be followed for a minimum of 4 years in order to capture the start, middle, and end of his or her own height velocity curve. This method, coupled with the 2- to 3-year population variability in the timing of the growth spurt, implies use of a longitudinal survey covering 6 to 7 years, clearly a radically different design from a mixed longitudinal survey consisting of two measurements 1 year apart.

Another important consideration when choosing among the various designs is cost. The costs of the designs depend on three factors, which can be summarized as recruitment, retention, and measurement. Recruitment involves identifying suitable subjects for the study and persuading them to take part. The corresponding cost per recruit depends on the time required, which in turn depends on the sampling fraction, i.e., the sample size required relative to the available population. Retention involves maintaining contact with previously recruited subjects and retaining highly trained and hence valuable staff in employment. Retention of subjects is important only for longitudinal studies. Measurement is the process of visiting and measuring children, and the cost of each measurement is the same whether the design is cross-sectional or longitudinal.

The relative costs of the cross-sectional and longitudinal designs depend on the relative costs of recruitment and retention. For a given number of child-measurement occasions, recruitment costs are minimized with a longitudinal design as the number of subjects is minimized, whereas retention costs in a longitudinal design are minimized if staff numbers are minimized. Elapsed time is a distinct resource that impacts directly on cost. A cross-sectional design is completed more quickly than a mixed longitudinal design, which is in turn more quickly completed than a longitudinal design.

Thus, the issues determining which design to use depend on the uses to which the reference is to be put, in particular the priority attached to assessing velocity (in addition to distance) and the time period over which to measure it; and the time and cost resources likely to be available for data collection.

A document produced for the MGRS discusses these issues in more detail and concludes that generally, and more particularly when recruitment costs are high, a mixed longitudinal design is preferable to a cross-sectional design.* A longitudinal design is much more expensive than a mixed design and is probably not suitable for the present purposes.

Sample-size derivation for number of populations and number of children within each population

The prescriptive form of growth standard proposed for use here is fundamentally different from the usual growth reference. The conventional approach is to obtain a sample of children representative of the country or region of interest, where the key issue is representativeness via randomization (subject to mild exclusion criteria). The proposed prescriptive growth standard involves identifying a sample of countries that is, in some sense, representative of all countries and then drawing samples of children from each country that are broadly representative of their country, but subject to more stringent inclusion and exclusion criteria to oversample those who have experienced unconstrained growth.

Thus, there are two distinct design questions to address: How many children per country should be sampled? How many countries should be sampled? The question "Should sample size vary from country to country?" might be justified if larger countries were to provide larger samples, e.g., with sample size weighted by population size, so that the sample reflected, to some degree, the global population. However, formal representativeness of the global population is not a requirement of the standard, so this question is not pursued further here. Instead it is argued that sample size should, by symmetry, be constant across countries. This means that the first question can be rephrased as "How many children overall?" The number of children divided by the number of countries gives the number of children per country.

* Cole TJ, Frongillo EA. Sampling scheme for children aged 18 to 71 months: the use of mixed longitudinal designs. WHO Multicentre Growth Reference Study internal report, 2000.

How many children overall?

At present, no satisfactory statistical basis is used for determining sample size in growth surveys. A common rule of thumb, which was used in the MGRS, is 50 or 100 or 200 subjects per age group. Each site planned to measure 70 subjects every 3 months longitudinally, and the same density of 70 measurements per 3 months defined the sample size for the 18- to 71-month cross-sectional survey. Increasing this number to 78 to cover for refusals and dropouts gave a total of $78 \times 54/3 = 1{,}400$ cross-sectional subjects per site. With a total of six sites, this gave an overall sample size of $78 \times 6 = 468$ longitudinally and $1{,}400 \times 6 = 8{,}400$ cross-sectionally. The latter number corresponds to 210 subjects of each sex per 3 months of age. (In practice, the numbers of subjects recruited were considerably larger than this, to protect against dropouts.)

The problem with this rule of thumb is that the width of the age group is not specified. Instead of 3 months, the measurement age could be 1 week or 1 year, which would imply a sample size either 13 times larger or 4 times smaller. Furthermore, cross-sectional surveys generally do not rely on recruitment to narrow age groups. Instead, children are recruited across a wide age band, and within the band the ages are assumed to be approximately uniformly distributed. Modern statistical techniques for constructing growth standards can deal with data at the exact age of measurement and do not require them to be grouped.

The assumption at present is that the data are analyzed in age groups. Within each group, the measurement distribution can be summarized as the mean μ and standard deviation σ, and assuming a normal distribution, each required centile is calculated as $\mu + z\sigma$, where z is the z-score (or normal equivalent deviate) corresponding to the particular centile. The precision of the centile estimate depends on three things: the precision of the mean μ, the precision of the standard deviation σ, and z. These are combined in the following formula:[8]

$$SE(Centile) = \sigma \sqrt{\frac{1 + \frac{1}{2}z^2}{n}} \qquad (1)$$

where SE is the standard error and n the sample size. This shows that the most precise centile is the mean where $z = 0$ (which is also the median for normally distributed data). Here the standard error is $\sigma/\sqrt{n}$. Extreme centiles are less precise; for example, the standard error for the 2nd centile, where $z = -2$, is $\sqrt{3}$ times or 70% larger than for the median.

Alternatively, the standard error can be expressed as a percentage of the mean μ, as follows:

$$\%SE(Centile) = 100\frac{\sigma}{\mu} \sqrt{\frac{1 + \frac{1}{2}z^2}{n}} \qquad (2)$$

Here the %SE for the mean is $100\sigma/\mu/\sqrt{n}$.

Equation (1) shows that the precision of the centile depends on the ratio $\sigma/\sqrt{n}$, where σ varies with age. This ratio can be made constant across age groups by ensuring that n is proportional to σ^2. Thus, at ages at which the variability is increased, notably in puberty, n needs to be increased appropriately to compensate [9].

Alternatively, if the percentage error is to be kept constant, as in equation (2), which is in many respects a better strategy, then the key ratio is $\sigma/\mu/\sqrt{n}$, where σ/μ is the population coefficient of variation (CV) or proportional standard deviation. For weight and height, the CV, like the SD, peaks in puberty, and for a constant %SE the sample size should be chosen to be proportional to CV^2. As an example, growth references based on the LMS method (discussed later) estimate the population CV as the quantity S. In the British 1990 reference, the S value for weight at the age of 5 years was 0.12 (12%) in boys and 13% in girls, rising to 18% in boys at 14 and 19% in girls at 11 years. With sample size proportional to S^2, this implies sampling more than twice as many subjects at the peak of puberty than at age 5.

Two other important features of growth standard data also need to be considered. The first is the smoothing that takes place across age. The means and standard deviations μ and σ by age group are plotted against mean age, and smooth curves are drawn through them. This process "borrows strength" from neighboring age groups, so that the standard errors of μ and σ for a particular age group are shrunk and the centile standard errors are smaller than predicted by equation (1). This increase in precision can be thought of as an effective increase in sample size, by up to three times [10].

The standard error is greatest at the extremes of age, and it can be reduced in two ways: either by oversampling the youngest and oldest age groups, or by sampling outside the intended age range. For example, if the reference is to cover the age range from 5 to 20 years, then including data from 4 to 22 years will make the centile estimates at the ages of 5 and 20 much more precise.

The discussion so far has also ignored the possible longitudinal nature of the data. Assume that the study uses a mixed longitudinal design in which some subjects are measured more than once, and that one purpose of the reference is to estimate velocity v. This is defined as the rate of change of the measurement, so $v = (b - a)/\Delta t$ where a and b are the two measurements and Δt is the time interval between them. Assume that a and b have the same variance σ^2, then the standard error of mean velocity is given by

$$SE(\bar{v}) = \frac{\sigma}{\Delta t} \sqrt{\frac{2(1 - r)}{n}} \qquad (3)$$

where n is the sample size and r the correlation between a and b. The standard error is inversely related to $\sigma/\sqrt{n}$, as before, but now it also depends on the correlation

coefficient r and the time interval Δt. To achieve a given precision for the velocity, the correlation between measurements needs to be taken into account. For example, the correlation between heights measured 1 year apart exceeds 0.98 at 5 years [11] but is only about 0.9 during puberty. This implies that for children during puberty, a sample size n more than four times greater than that for children at age 5 would be required to ensure the same standard error for the mean annual velocity.

The dependence of the velocity precision on Δt in equation (3) has design implications. If velocity measurements from different subjects are to be equally informative, the value of equation (3) needs to be constant from child to child. This means either that Δt should be constant, i.e., that all children are measured at exactly the same ages (not easy to do in practice), or that individual estimates of velocity should be weighted inversely as their variance, i.e., by Δt^2, to compensate for differences in Δt.

To summarize, the precision of distance centiles involves the size of the group and the population variance, whereas for velocity centiles the age-on-age correlation and time interval between measurements are also important. Because of the extra variability in growth at puberty, the sample size requirements to achieve a given size of standard error are considerably increased at this time, by a factor of four or more. Peak height velocity (PHV) occurs at about the age of 12 in girls and 14 in boys, so in addition to analyzing the sexes separately, which is clearly necessary, it may also be optimal to have different age profiles for the sample size by sex.

This is a simple summary of a complex situation. Cole and Frongillo* have considered it in more detail, covering acceleration in addition to velocity and distance, and they conclude that the mixed longitudinal design is a good compromise in terms of precision for estimating distance and velocity.

The focus so far has been on distance in narrow age groups or velocity over short periods of time. In practice, though, the fitting of growth standards involves estimating smooth curves that cover the whole age range of the study. The LMS (lambda-mu-sigma) method, for example, is used to construct distance centiles, and it estimates the age-changing distribution of the measurement in terms of its median M, coefficient of variation S, and skewness L (Box-Cox power to transform to normality), each represented as a cubic smoothing spline curve plotted against age estimated by penalized maximum likelihood [12]. The LMS method is an extension of the normal distribution model of equation (1) to a model that includes an adjustment for skewness. This is valuable for measurements such as weight and body mass index (BMI) where the distribution is markedly skewed.

The complexity of the shape of each estimated curve in the LMS method depends on the particular measurement and the age range of the data, so that median height from birth to adulthood is a more complex curve shape than, for instance, median skinfold thickness from 5 to 10 years. The greater the complexity of the curve, the greater the sample size required to estimate it with adequate precision. Pan and Cole [13] recently considered issues of curve complexity when fitting the LMS method. They used as their example data from the Third Nationwide Dutch Growth Survey, a sample of 20,000 heights and weights for each sex from birth to 20 years [14], but they also considered a random 10% subsample of 2,000 for comparison purposes. This corresponds to a sample density of 250 for each sex per 3 months reduced to just 25 for each sex per 3 months (as against the nominal figure of 78 for each sex per 3 months used in the MGRS). They found that reducing the sample size by a factor of 10 made remarkably little difference to the estimated L, M, and S curves, implying that the original survey had been larger than necessary and hence overpowered. Bearing in mind that the age range being discussed for the international growth standard for preadolescent and adolescent children is narrower than the Dutch survey's range of 0 to 20 years, the implication is that a total sample size of less than 2,000 for each sex should be sufficient.

How many countries?

The choice of the number of countries depends, like the choice of the number of children, on the research questions to be asked and the resources available. The possible number of countries to include varies from one (which equates to a conventional national reference) to all countries (approximately 200). Other things being equal, the total cost rises steeply with the number of countries included, an argument for a fairly small number of countries. On the other hand, if the aim is to quantify intercountry differences, then a relatively large number of countries is needed. In practice, the political imperative is to include sufficient countries for the standard to claim international representativeness, which can probably be achieved with fewer than 10 suitably chosen countries.

A statistical case can be made to justify a given number of countries, predicated on the precision of the intercountry differences. But this implies that the aim of the study is to estimate intercountry differences, and in the MGRS this was not the case. The assumption has been that, by sampling children of high socioeconomic status in each country, the children can be made relatively homogeneous across countries and the intercountry differences minimal, obviating the need to make formal comparisons.

* Cole TJ, Frongillo EA. Sampling scheme for children aged 18 to 71 months: the use of mixed longitudinal designs. WHO Multicentre Growth Reference Study internal report, 2000.

Ultimately, the number of countries to sample is a political rather than a statistical issue. But a number between 5 and 10 is likely to prove suitable.

Growth parameters of height, weight, and other measurements

The cross-sectional age-related frequency distribution of anthropometric characteristics (including height, weight, BMI, and skinfolds, among others) can be estimated by using techniques like the LMS method [12] or the EN (exponential-normal) method [15], which summarize the distribution in terms of age-changing curves representing the median, coefficient of variation, and transformation required to adjust for skewness. These methods offer several advantages over previous statistical approaches to growth curve construction:

» They summarize the distribution in terms of its age-related moments, the mean, variance, skewness, etc., which apply across the age range and so "borrow strength" from nearby data;
» The incorporation of a skewness adjustment allows for measurements such as BMI or skinfolds to be modeled, which in the past posed problems because of their skew distributions;
» The LMS method uses cubic smoothing splines for the curve-fitting, which are simple to fit and yet allow for considerable complexity in curve shape;
» The moment curves allow any required centile curves to be calculated, which can be extrapolated into the tails of the distribution, providing, for example, curves as extreme as –3 or +3 z-scores, corresponding to the 0.1th and 99.9th centiles;
» Dedicated software exists for fitting the LMS method, for both statistical specialists and nonspecialists.

The foregoing applies mainly to distance standards, i.e., based on cross-sectional data, and it can be applied equally to velocity standards based on longitudinal data. However, conventional velocity standards suffer from two important problems that are not widely acknowledged:

» Using the velocity standard more than doubles the amount of work required to assess an individual's growth curve. Two charts, one for velocity and one for distance, are needed rather than one, and the velocities need to be calculated from the individual measurements to plot them on the velocity chart;
» Velocity needs to be adjusted for regression to the mean; on average, a subject's second measurement is less extreme than the first (it has "regressed toward the mean") so that small children grow faster than large children on average [16]. The assessment needs to take this into account.

There is an alternative approach to the construction of velocity standards that directly addresses these concerns. It exploits the idea of velocity as "centile crossing" on the chart, where normal or median growth is seen as tracking along the centiles, whereas growing faster or slower than median growth is upward or downward centile crossing. A given rate of centile crossing then corresponds to a particular velocity centile, and this can be represented on the chart by a set of "thrive" lines, analogous to centile lines, which demonstrate the required degree of centile crossing. A child whose growth curve tracks along these thrive lines (so called because they test for failure to thrive) is then growing on the specified velocity centile [17, 18].

The thrive-line approach to growth velocity avoids the need for a velocity chart, since the thrive-line information is provided on a transparent overlay to place on the distance chart. The thrive lines correspond to a particular centile over a specified time period, e.g., the 5th weight velocity centile over a 4-week period [18]. The thrive lines also incorporate an adjustment for regression to the mean, which addresses the second of the two concerns above. The algebra underlying the thrive lines is actually very simple, and they depend only on the correlation structure between measurements at a series of ages [18].

The choice of anthropometric characteristics should obviously include height, weight, and hence BMI. Skinfolds provide information on fat content and fat distribution, but waist circumference is probably a better source of such information because it is easier to measure. Another approach to consider is the measuring of bioelectrical impedance, which is a proxy for lean mass and thus can be used to adjust BMI for lean mass, providing a more direct measure of fat mass than BMI itself [19].

Modeling of the adolescent growth spurt

The adolescent growth spurt is seen most clearly with height and weight. The growth velocity in individual children rises from a low point just before adolescence to a peak at a mean age of approximately 12 in girls and 2 years later in boys, then falls equally rapidly to zero as adulthood is reached. The timing of this growth spurt varies from child to child, with a standard deviation of about 1 year, reflecting the individual tempo of growth, with the result that data collected cross-sectionally blur the form of the growth spurt and flatten the median curve, as described originally by Merrell [20]. For this reason, Tanner produced a tempo-conditional reference that adjusted for the flattening of the median curve by representing it as the growth pattern of a child of average height, average age at PHV, and average PHV [6, 7]. There is some debate about the added value of the tempo-conditional reference as compared with a cross-sectional reference [21].

At the individual rather than the population level, the adolescent growth spurt can be estimated by fitting either a parametric or a semiparametric model to individual growth curves. Of the various parametric models

available, the most suitable for the age range from 5 to 20 years is the Preece–Baines model 1 [22], which summarizes the growth curve in five parameters that are closely related to the biological parameters of age and height at takeoff, age and height at PHV, and adult height. It is important to note that the Preece–Baines model, like other parametric models, requires adult height to be measured for each individual to ensure a good fit to the growth curve.

An alternative is the semiparametric approach, where each individual curve is fitted as a cubic smoothing spline. The age at PHV for the individual is then obtained as the age when the first derivative of the spline curve, i.e., the height velocity curve, is at a maximum. This is a more flexible approach than the parametric model, in that a spline curve of up to 7 equivalent degrees of freedom (an indication of curve complexity) is adequate to represent all the different shapes of growth curve likely to be seen during puberty, and its first derivative follows closely the shape of the height velocity curve, particularly its peak. In addition, there is no requirement for adult height to be known.

Other modeling issues

There are two other modeling issues that need to be borne in mind, given that the new survey is intended to extend the WHO MGRS. The first issue is the need for the centiles in the two surveys to match at the age where they meet. The NCHS reference consisted of an early component based on Fels data and a later component based on National Health Examination Surveys [23]. The centiles for the two components did not match, and as a result the NCHS centiles, and the prevalences of malnutrition based on the NCHS centiles, showed a disjunction between 2 and 3 years [2]. It is very important that the model-fitting process avoid this happening with the MGRS and the International Growth Standard for Preadolescent and Adolescent Children. The best way to achieve it would be to add MGRS data for the older children to the international growth standard for preadolescent and adolescent children dataset to achieve a smooth transition between the two.

The second and related issue is the secular trend in obesity, which means that during the time interval between data collection in the two surveys, children will have become appreciably more obese. Recent data for UK children aged 5 to 10 years suggest an increase in overweight prevalence of about 1% per year between 1994 and 2003 [24], which would translate to a difference in prevalence of 5% or more between the MGRS and the international growth standard for preadolescent and adolescent children data. Thus, the issue of a disjunction between the two surveys at age 5 is of paramount concern, since the centiles for weight and particularly BMI in the two surveys are likely to be offset to this extent.

References

1. Garza C, De Onis M. A new international growth reference for young children. Am J Clin Nutr 1999;70: 169S–72S.
2. Ogden CL, Kuczmarski RJ, Flegal KM, Mei Z, Guo S, Wei R, Grummer-Strawn LM, Curtin LR, Roche AF, Johnson CL. Centers for Disease Control and Prevention 2000 growth charts for the United States: improvements to the 1977 National Center for Health Statistics version. Pediatrics 2002;109:45–60.
3. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. Stat Med 1998;17:407–29.
4. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in the Netherlands 1955–1997. Pediatr Res 2000;47:316–23.
5. Sempé M, Pédron G, Roy-Pernot M. Auxologie: méthode et séquences. Paris: Theraplix, 1979.
6. Tanner JM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. I. Arch Dis Child 1966;41:454–71.
7. Tanner JM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. II. Arch Dis Child 1966;41:613–35.
8. Healy MJ. Notes on the statistics of growth standards. Ann Hum Biol 1974;1:41–6.
9. Goldstein H. Sampling for growth studies. In: Falkner F, Tanner JM, eds. Human growth: a comprehensive treatise, 2nd ed. New York: Plenum Press, 1986:59–78.
10. Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr 1990;44:45–60.
11. Cole TJ. Growth monitoring with the British 1990 growth reference. Arch Dis Child 1997;76:47–9.
12. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305-19.
13. Pan H, Cole TJ. A comparison of goodness of fit tests for age-related reference ranges. Stat Med 2004;23: 1749–65.
14. Roede MJ, Van Wieringen JC. Growth diagrams 1980. Netherlands third nation-wide survey. Tijdschr Soc Gezondheidsz 1985;63(suppl):1–34.
15. Royston P, Wright EM. A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. J R Stat Soc [Ser A] 1998;161:79–101.

16. Cameron N, Preece MA, Cole TJ. Catch-up growth or regression to the mean? Recovery from stunting revisited. Am J Hum Biol 2005;17:412–7.
17. Cole TJ. 3-in-1 weight-monitoring chart. Lancet 1997;349(9045):102–3.
18. Cole TJ. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. Stat Med 1998;17:2697–707.
19. Wells JCK, Williams JE, Fewtrell M, Singhal A, Cole TJ. A simplified approach to analysing bio-electrical impedance data in epidemiological surveys. Int J Obes 2006;(in press).
20. Merrell M. The relationship of individual growth to average growth. Hum Biol 1931;3:37–70.
21. Wright CM, Booth IW, Buckler JM, Cameron N, Cole TJ, Healy MJ, Hulse JA, Preece MA, Reilly JJ, Williams AF. Growth reference charts for use in the United Kingdom. Arch Dis Child 2002;86:11–4.
22. Preece MA, Baines MJ. A new family of mathematical models describing the human growth curve. Ann Hum Biol 1978;5:1–24.
23. Hamill PVV, Drizd TA, Johnson CL, Reed RB, Roche AF. NCHS growth curves for children birth–18 years. Washington, DC: National Center for Health Statistics, 1977.
24. Stamatakis E, Primatesta P, Chinn S, Rona R, Falascheti E. Overweight and obesity trends from 1974 to 2003 in English children: What is the role of socio-economic factors? Arch Dis Child 2005;90:999–1004.

Indicators of biological maturation and secular changes in biological maturation

Gaston P. Beunen, Alan D. Rogol, and Robert M. Malina

Abstract

Commonly used indicators of biological maturation are discussed, including sexual, skeletal, morphological, and dental maturity, and the hypothalamus–pituitary–end organ axes that regulate the growth and maturation processes. Interrelationships among indicators and the tempo, timing, and sequence of maturational events are also considered. Environmental factors that influence the level of maturity at a given point in time and the process of maturation are also discussed: undernutrition, obesity, ethnic/racial background, social class, familial characteristics, climate, and altitude. Recommendations for the design of studies of maturational events are made, and an overview of secular changes before and after 1970 is provided. The review concludes with specific recommendations for the inclusion of a maturity indicator or maturity indicators in the construction of an international growth standard for preadolescent and adolescent children

Key words: Growth reference, hormones, maturity, secular change

The concept of biological maturation

Maturation is a process that marks progress toward the adult (mature) state. Maturation is a process, whereas maturity is a state. All tissues, organs, and organ systems of the body mature, but they do so at different times and rates. As a result, assessment of biological maturity status varies with the bodily system considered. Of necessity, therefore, the concept of maturation is operational. The more commonly used systems for the assessment of maturation are the skeletal, reproductive (sexual), and somatic systems; hence, the terms skeletal, sexual, and somatic maturation are standard in the growth literature. Dental maturation (eruption and calcification) is occasionally used, but it tends to proceed independently of the other three systems. Biochemical and hormonal maturation, as steering mechanisms for the other systems, must also be considered.

Maturation of different systems tends to proceed independently of chronological (calendar) age, so that chronological age is not a good indicator of biological maturity. Nevertheless, the growth and maturity status of an individual or sample of individuals is routinely placed in the context of chronological age.

In constructing objective, reliable, and valid indicators of biological maturity status, it is of importance that the indicators reflect the maturation of a biological system, occur in all individuals as they progress toward the adult state, and reach the same endpoint, i.e., the mature or adult state. The indicators should also, to some extent, be independent of growth (size attained), i.e., they should not quantify the growth status of a tissue, an organ, or a biological system. Finally, a relevant indicator should be applicable throughout the entire maturation process, but the reality of variation among systems precludes this criterion [1–6].

Indicators of biological maturation

Skeletal maturity

The maturation of the skeleton is widely recognized as the best single indicator of maturity status [6]. All children start with a skeleton of cartilage and progress

Gaston P. Beunen is affiliated with the Department of Biomedical Kinesiology, Faculty of Kinesiology and Rehabilitation Sciences, Katholieke Universiteit Leuven, Leuven, Belgium; Alan D. Rogol is affiliated with the Department of Pediatrics, University of Virginia, Charlottesville, Virginia, USA; Robert M. Malina is affiliated with Tarleton State University, Stephenville, Texas, USA.

Please direct queries to the corresponding author: Gaston P. Beunen, Department of Biomedical Kinesiology, Faculty of Kinesiology and Rehabilitation Sciences, K.U. Leuven, Tervuurse vest 101, B3001, Leruven, Belgium; e-mail: gaston.beunen@faber.kuleuven.be.

toward the fully ossified, adult axial skeleton. In the case of the tubular bones (long and short bones), maturity is attained when the epiphyses are fused with their corresponding diaphyses; in the case of round or irregularly shaped bones, maturity is defined by adult morphology (shape). The bones comprising the craniofacial skeleton differ in embryonic origin, and their growth and maturation are approached differently. They are not considered in this discussion of skeletal maturation.

The bones of the hand and wrist provide the primary basis for assessing the maturity status of the child, although the knee, hip, and foot have also been used. The progress of maturation of the skeleton is ordinarily monitored with standardized radiographs, and assessment of maturity is based on changes occurring from initial ossification to adult morphology of individual bones. Criteria for individual bones are characterized as maturity indicators—specific features of individual bones that are universal and occur regularly in a definite, irreversible order. Three methods for the assessment of skeletal maturity—the Greulich-Pyle, Tanner-Whitehouse, and Fels methods—are commonly used at present.

The *Greulich–Pyle method* [7] is based on the original work of Todd [8], and it is sometimes called the atlas or the inspectional technique. The atlas consists of sex-specific radiographs representative of the maturity status at a given chronological age from birth to 19 years. The radiograph that was most typical of about 100 radiographs of each sex at each age level was selected as the reference plate. Each bone on the standard plates represents its median appearance at a given chronological age (however, in some plates of the atlas, there is considerably variation for a given chronological age). The method is based on the radiographs of a large sample of children from the Brush Foundation Study. The children were from families of high socioeconomic status in Cleveland, Ohio, USA.

The skeletal maturity of a child is determined by comparing his or her hand-wrist skeleton to the standard plates of the atlas. Skeletal maturity is expressed as a skeletal age. There is, however, variation in how the method is applied. Quite often, the assigned skeletal age is that of the plate which most closely matches that of the child. This overlooks variation among bones in the hand-wrist and also variation among standard plates. More appropriately, the method should be applied by matching each individual bone to the atlas plates. Accordingly, the skeletal age of the plate with which the individual bone most closely coincides is noted, and the skeletal age assigned to the child is the median value of the skeletal ages of all bones [9].

The *Tanner–Whitehouse method* is sometimes called the bone-specific approach [10, 11]. Maturity indicators were defined and described for each bone. Each indicator is expressed as a stage from initial ossification to union (radius, ulna, metacarpals, phalanges) or adult morphology (carpals), and a point score is assigned to each stage. Twenty bones are used: the radius, ulna, seven carpals (excluding the pisiform), and the metacarpals and phalanges of the first, third, and fifth digits (rays). The scores are summed and can be expressed either as a maturity score or as a skeletal age. The maturity scale (0 to 1,000) was constructed to minimize the overall disagreement between the results from the long and the round bones.

The first version of the method (Tanner–Whitehouse I) [10, 11] provided a skeletal age based on the sum of maturity scores for 20 bones. The second version (Tanner–Whitehouse II) [12] provided three different scales and skeletal ages: a 20-bone scale, an RUS (radius, ulna, short bones) score (13 bones), and a CARP scale for the seven carpal or round bones. Both the Tanner–Whitehouse I and the Tanner–Whitehouse II skeletal maturity references are based on a sample of about 3,000 healthy British children. In the second version, the final stage of a number of bones was no longer assessed, and the scoring system was modified, but the maturity indicators were not changed. The third version of the method (Tanner–Whitehouse III) [13] considers only the RUS and carpal bones and no longer includes a 20-bone skeletal age, and the reference values are now based on samples of British, Belgian, Italian, Spanish, Argentinean, US (a well-off sample from the suburbs of Houston, Texas), and Japanese children.

The *Fels method* is bone-specific and is based on a sample of middle-class children from south-central Ohio, USA, enrolled in the Fels Longitudinal Study [14]. The authors defined an extensive series of maturity indicators for all bones of the hand-wrist skeleton [14]; ratios between linear measurements of epiphyseal and diaphyseal widths for individual long bones were included among the indicators. The potential of each indicator was tested on its ability to differentiate between individual children of the same chronological age and on its universal appearance, reliability, validity, and completeness. The resulting Fels method is based upon the final grading of 85 grade maturity indicators for the radius, ulna, carpals, metacarpals, and phalanges, and 13 measured ratios of epiphyseal and diaphyseal diameters of the radius, ulna, metacarpals, and phalanges. The number of indicators to be assessed at a given chronological age varies with chronological age and sex and is relatively large at some ages; however, most indicators are assessed simply as present–absent or maximally on a five-grade scale. The chronological age and sex of the child and the ratings and ratios are entered into a microcomputer, which calculates a skeletal age and associated standard error of estimate.

Other methods for the assessment of skeletal maturity have been proposed. Some are of historical interest, and others are less commonly used [3, 6]. At present, several computer-based protocols have been applied to

the Tanner–Whitehouse II method, and the experimental results are reasonably consistent with the ratings of expert assessors [15, 16].

The three currently used methods for the assessment of skeletal maturity are similar in principle but differ in maturity indicators, scales of maturity (scores, skeletal age), and reference samples. The Greulich–Pyle and Fels methods provide a single skeletal age, whereas the Tanner-Whitehouse method provides several skeletal ages. A skeletal age corresponds to the level of skeletal maturity attained by a child relative to the reference sample for each method. Given differences in the methods and in the reference samples for each, skeletal ages derived from each are not equivalent. In fact, the skeletal maturity status of a child rated by all three methods may be quite different [1, 3–5, 17]. Regardless of the method used, quality control in assessment is essential. Variation within and between assessors can be considerable and should be reported.

Skeletal age has limited utility by itself. The utility of skeletal age as a maturity indicator is based on its relationship to a child's chronological age. Skeletal age may simply be compared with chronological age, may be expressed as the difference between skeletal age and chronological age (i.e., skeletal age minus chronological age), or may be expressed as a ratio of skeletal age to chronological age. There is considerable variation in skeletal age at each chronological age level. The standard deviations of the RUS bone age (Tanner–Whitehouse III) is approximately 1 year from the age of 5 years in both sexes to 14 years in girls and to 16 years in boys (Tanner et al. [13], p. 10).

The advantages of skeletal maturity as an indicator of biological maturity are several: it gives reasonably precise and reliable estimates, is applicable throughout the postnatal maturation period, and reflects maturation of an important biological system. Its disadvantages are that it involves exposure to low-level radiation, it requires training and quality control, and the stages (maturity indicators) are somewhat arbitrary and suggest discrete steps in a continuous process [1, 3–5].

Sexual maturity

Sexual maturation is a process that extends from the early embryonic differentiation of the sexual organs to full maturity of these organs and fertility. Puberty is a transitional period between childhood and adulthood during which the sex organs and the reproductive system mature and the growth spurt takes place. Major psychological, behavioral, cognitive, and emotional changes also occur during puberty. Individual differences in timing and tempo are considerable at this time.

The assessment of sexual maturation is based on secondary sex characteristics: breast development and age at menarche in girls, genital (penis and testes) development in boys, and pubic hair in both sexes. Development of the breasts, genitals, and pubic hair is most often rated on five-point scales described by Tanner [5]. The stages should not be identified as "Tanner stages" but as stages of sexual maturation with identification of the specific characteristic(s) (breast, pubic hair, or genitals) assessed. The stages of each characteristic are neither equivalent nor interchangeable. Stage 1 of each characteristic indicates the prepubertal state (absence of development) and stage 2 the initial, overt development of each characteristic that marks the transition into puberty. Stages 3 and 4 mark progress in maturation, and stage five 5 indicates the adult (mature) state.

Ratings of stages of secondary sex characteristics are ordinarily made by individual observation at clinical examination. Sometimes, as in the Harpenden Growth Study [18, 19], the examination was made from standardized, nude photographs. In nonmedical settings, self-assessments by youths are increasingly used. Self-assessments should be done privately in a quiet room using good-quality photographs of the stages and simplified descriptions. There is obviously a need for quality control (intra- and interobserver reliability), and in the case of self-assessment concordance with experienced assessors should be verified. Overall reproducibility by experienced assessors is generally good, with about 80% of agreement in assigning the stages, although some studies report a percentage of agreement as low as 40% [3].

Age at menarche, the first menstruation, is perhaps the most widely monitored secondary sex characteristic in females. It can be obtained in three different ways: prospectively (longitudinal design), by interrogating the same girls at regular intervals of 3 to 6 months; retrospectively, by interrogating postmenarcheal girls or women and asking them to recall when they experienced their first menstruation; and status quo, by interrogating large samples of girls approximately 9 to 16 years of age about their menarcheal status (i.e., pre- or postmenarcheal, see below). The first two methods provide ages at menarche for individuals, whereas the status quo method provides an estimated age at menarche for a sample and does not apply to individuals.

Other secondary sex characteristics include axillary hair in both sexes and facial hair and voice change in boys. As a rule, these are late-developing indicators during puberty and are not widely used in studies of biological maturation. A more direct estimate of genital maturity in boys is provided by testicular volume. The method is used primarily in the clinical setting and requires a series of ellipsoid models of known volume, which have the shape of the testes (Prader orchidometer) [20, 21]. The models range in volume from 1 to 25 ml; a volume above 4 ml marks the beginning of puberty.

The ages at which specific stages of sexual maturity are reached are ordinarily derived from longitudinal studies in which children are examined at regular intervals, preferably every 3 months, starting in late childhood (prepuberty) and continuing through puberty into early adulthood. Data obtained from prospective studies provide estimates of the age at initiation of a stage and duration of a stage. Mean ages and associated standard deviations can be calculated. Such longitudinal studies require, of course, long examination periods and are most often restricted in sample size and representativeness of the sample. Cross-sectional designs (status quo) provide ages of "being in a particular stage." Two pieces of information are needed: the exact chronological age of the child and whether or not the child is in a particular stage of sexual maturation or, in the case of girls, pre- or postmenarcheal. The percentages of children in a particular stage at each age are used with probits or logits to obtain sample statistics (median, means, and standard deviations) for each stage of a characteristic or for age at menarche. The percentages of individuals in each stage of a secondary sex characteristic increase with chronological age, and the maturity curves have a sigmoid shape.

Secondary sex characteristics are reasonably easy to determine, reflect an important biological system, and are closely related to underlying hormonal axes. On the other hand, secondary sex characteristics have limitations, in that the stages are somewhat arbitrary and discrete, they are limited to puberty, and the method of assessment is invasive in nonclinical settings (not necessarily true for self-assessment). Moreover, the use of secondary sex characteristics may have associated sanctions among some cultural groups.

Biochemical and hormonal maturity

Growth and adolescent maturation surely depend on specific hypothalamic–pituitary–end organ axes. The process of fetal growth does not depend very importantly on the fetal hypothalamic–pituitary function; however, the process of fetal differentiation does.

Hypothalamic–pituitary–thyroid axis

The physiological maturation of the thyroid is apparent as early as the 8th week of gestation [22]. By the 10th to the 11th week, iodine trapping and synthesis of thyroid hormones occur. Until birth the metabolically inactive reverse triiodothyronine (rT_3) predominates, only to be followed by a large burst of thyrotropin (TSH) secretion just after birth and a switch to the more metabolically active T_3 by a specific deiodinase enzyme. There are only slight differences in the normal thyroid axis hormone levels in the first year or two of life compared with levels in the adult. The hormone levels then remain virtually the same until puberty, when estrogen raises thyroxin-binding globulin (TBG) levels. Although thyroid hormones are not responsible for the pubertal growth spurt or sexual maturation, they are thought to be permissive for these processes. Adequate thyroxin is necessary for normal growth in infancy and childhood and also for growth hormone (GH) gene expression, and thyroxin may also act directly on cartilage [23].

Hypothalamic–pituitary–adrenal axis

The hypothalamic–pituitary–adrenal axis shows hormonal activity beginning between the 8th and 12th weeks of gestation. Corticotropin-releasing hormone (CRH) from the hypothalamus regulates the growth of pituitary corticotrophs, adrenocortical differentiation, and steroidogenic maturation of the fetal hypothalamic–pituitary–adrenal axis. The adrenal gland at birth is composed mainly of the definitive (mineralocorticoid) and the very much larger fetal dehydroepiandrosterone (DHEA) zones. As the child matures, the adrenal gland forms a focal reticular and then a continuous reticular zone. It is this zone that makes adrenal androgens under the stimulus of corticotropin and perhaps other adrenal androgen-stimulating hormones. The process of *adrenarche* marks the transition of this zone as it releases greater and greater quantities of the adrenal androgens, DHEA and its sulfate (DHEA-S), and androstenedione, precursors of both more potent androgens (testosterone) and estradiol. There is a steep rise, perhaps 4- to 50-fold, in DHEA-S and androstenedione secretion. Adrenarche usually occurs at the same time as the mid-childhood growth spurt and together with the preadolescent fat spurt. This process is independent of gonadotropin-induced "true" puberty. However the mid-childhood growth spurt is of much less magnitude than the pubertal growth spurt (see below) and is quite variable in its timing, tempo, and magnitude, depending on the state of pubertal gonadal development. It is not a useful parameter for linking linear growth to the "biochemical" (e.g., hormonal) measurements.

Hypothalamic–pituitary–gonadal axis

Sexual determination (testicular development) occurs at conception. Sexual differentiation (genital development) is the process by which the manifestations of that determination become overt. Male sexual differentiation requires the expression of the product of the sex-determining region on the Y chromosome (SRY) to select the pathway that the bipotential gonad containing the Wolffian and Mullerian ducts and the external genitalia will take. The embryonic gonad differentiates along one or the other pathway beginning at approximately the sixth week of gestation under the influence of gene products of the sex chromosome and autosomes. Mutations in a number of transcription factors, for example, SRY, SOX9, and SF-1, may affect testicular determination [24]. Sexual differentiation

continues with stimulation of the Wolffian ducts and regression of the Mullerian structures in boys. The former are stimulated directly by testosterone to form the vas deferens, epididymis, and seminal vesicles. Testosterone also potentiates the effects of anti-Mullerian hormone (AMH), also known as Mullerian inhibiting substance (MIS), to permit complete regression of these structures.

The male external genitalia require dihydrotestosterone (DHT) for full development. If this does not occur, the labial scrotal folds do not fuse completely and there is not an intact penile urethra.

Defects in testosterone production cause undervirilization of 46, XY infants. Before 10 weeks of gestational age, very little androgen production occurs. The critical period for androgen production (and action) occurs between the 10th and 20th fetal weeks, when the Leydig cell is dependent upon stimulation by luteinizing hormone (LH) and human chorionic gonadotropin (hCG). In the male fetus, the lack of full production of testosterone may lead to a genital phenotype ranging anywhere from that of a normal female to that of an incompletely developed male, with microphallus, scrotal hypoplasia, and undescended testes

In the first few days following delivery, the initially "high" levels of testosterone decline, only to rise again to approximately 8 nmol/L (230 ng/dL) sometime between weeks 3 and 12 [25]. These levels may be important for further alteration in genital development (for example, the priming of androgen target tissues for subsequent androgen-mediated growth and maturation) and/or brain development (for example, permanent virilization of the hypothalamus so that it secretes LH tonically, rather than cyclically as in the female [26]).

During the quiescent period between the neonatal–early infancy surge and pubertal development (the so-called *prepubertal hiatus* or *juvenile pause*), the full complement of structures and pathways for androgen synthesis, secretion, and action are present but are active at a very low level. Disorders of advanced puberty, e.g., central precocious puberty or peripheral "pseudo" puberty, may occur during this phase.

At puberty the levels of testosterone rise exponentially as the hypothalamic–pituitary–gonadal axis regains the active state. At first there are only small LH pulses, which cause the testis to produce small, but measurable, levels of testosterone. Since the negative feedback control system is operative at the (nearly) prepubertal, very sensitive range, these low levels of testosterone are capable of reducing gonadotropin-releasing hormone (GnRH) secretion and thus reducing LH release. As the boy matures, the GnRH pulse generator operates more like the adult generator, and the low, but rising, levels of testosterone are no longer able to have such exquisite negative feedback control. The sum of these two processes is increasing testosterone production, at first only at night (with the first pulse early in the first episode of deep sleep), and then into the day, but with a very distinct variation between day (early morning) and night, which may be as high as 10-fold. With "complete" maturation, there are fluctuations in testosterone concentration (perhaps 40%) during the 24 hours and a small diurnal variation, with the highest levels in the early morning.

Hypothalamic–pituitary–GH–IGF-I axis

GH is synthesized and secreted by 8 to 10 weeks of gestation, peaks at mid-trimester, and then decreases until delivery. The growth of the fetus is not particularly sensitive to the GH–insulin-like growth factor I (IGF-I) axis, since congenitally hypopituitary children have only a minor decrease in birth length. At birth the axis is quite active, with pulsatile GH release at relatively high amplitude. Throughout infancy and childhood, GH and its stimulation of IGF-I production are responsible (with adequate thyroid hormone levels) for the relatively constant growth rate. At puberty there is a marked increase (approximately 2.5- to 3.5-fold) in GH and IGF-I production, secondary to the estrogen-induced increase in pulsatile GH release. The levels of IGF-I may be 5- to 10-fold those of younger children and adults, especially during the period around PHV. The levels of GH (mean 24-hour production) and IGF-I peak coincidentally with peak height velocity (PHV) [27]. The variability in the release of these hormones precludes a simple relationship of their individual levels with height velocity; however, the mean levels over 24 hours correlate reasonably well, but not so tightly as to predict the attainment, timing, or tempo of PHV or of adult height.

Changing hormonal levels provide direct evidence of the maturation of specific structures and tissues that underlie the overt manifestations of biological maturation that are commonly assessed in growth studies, i.e., skeletal age, secondary sex characteristics, and adolescent growth spurt (see below). However, most of the hormones directly related to maturation are produced in a pulsatile manner, so that serial blood samples taken over relatively long periods (e.g., 8 or 24 hours) are required to adequately evaluate the hypothalamic-pituitary-end organ axes. For example, it is the increase in the pulse amplitude of GnRH that permits the increase in LH that drives the increase in testosterone and estradiol at puberty. Moreover, the collection of blood samples and associated assays require specialized equipment that precludes their use in large-scale surveys. Static levels of the steroid hormones may be measured in saliva or blood samples and may serve as "anchors" for several of the stages of adolescent development. The more recent third- and fourth-generation gonadotropin assays may permit the distinction of hypogonadotropic individuals from those who are normal, but prepubertal.

Somatic or morphological maturity

Body size by itself is not a valid indicator of biological maturity, since the adult state is not the same for all individuals. As such, it is not appropriate for use as an indicator of biological maturation. Concepts such as height age, i.e., the corresponding chronological age at which, in a population, a specific stature is on average attained, are not useful maturity estimates.

If longitudinal height data that span late childhood through adolescence are available, the characteristics of the adolescent growth spurt can provide two indicators of somatic maturity: age at the onset of the growth spurt in height (first inflection point of the adolescent growth curve, takeoff) and age at maximum velocity (second inflection point of the adolescent growth curve, PHV). Corresponding parameters of the growth spurt can also be derived for other linear measurements, e.g., sitting height and leg length.

If adult height is available (as in longitudinal studies), the percentage of adult height attained at a given age or the age at which a certain percentage of adult height is attained can be used as a maturity indicator. To accurately estimate the parameters of the growth curve, careful measurements that span adolescence and that are taken at regular intervals, at least two times per year (preferably three or four times a year), are needed. Curve-fitting techniques based on structural and nonstructural models have facilitated estimation of the parameters [28–30].

Structural models have a preselected form of the growth curve, and the mathematical parameters of the model have a predetermined biological meaning. Nonstructural models do not have a predetermined form, and the parameters may not be easy to interpret biologically.

The assessment of somatic maturity based on the parameters of the growth curve (age at onset and age at maximum velocity) is limited to the adolescent period, and only one or two biological events are considered. As noted, their derivation requires longitudinal measurements of individual children over a relatively large age span, but they do provide an accurate estimate for a major event in the pubertal period.

Percentage of adult height is calculated from present height and adult height. Adult height is measured if children are followed until adult stature is attained or can be estimated. Prediction formulas are available for European and American samples but have not been validated on other populations [9, 11–13, 31–33]. Attempts have also been made to predict adult stature without skeletal age [34, 35].

Use of the percentage of adult height as an indicator of somatic maturity is an indirect technique that requires the estimation of skeletal maturation, at least for the most accurate systems. It can, however, be applied throughout most of the maturation period, beginning in childhood, and reflects the progress toward maturity of an important biological characteristic.

Dental maturity

Dental maturity has been traditionally estimated from the ages of eruption of the deciduous and/or permanent teeth, the number of teeth present at a certain chronological age, or the age at which a specific number of teeth has erupted [36]. Eruption is only one event in the calcification process of teeth and has limited biological meaning. Moreover, the criteria for eruption (e.g., initial piercing of the gum line to complete eruption) vary.

Dental calcification, as evaluated on radiographs, also provides an indication of maturity status. Demirjian et al. [37] developed a scale of dental maturity based on the principles of the Tanner-Whitehouse [10] method for the assessment of skeletal age. The procedure requires panoramic radiographs of the seven teeth in one quadrant of the mouth (two incisors, the cuspid, two premolars, and the first and second molars). As in the Tanner–Whitehouse system, specific maturity indicators are identified for each tooth, the stages are scored on a maturity scale for each tooth, and the scores are subsequently summed to provide an overall dental maturity score.

Eruption and calcification of the teeth reflect the maturation of the dentition. Deciduous teeth erupt between about 6 and 30 months, and permanent teeth (excluding the third molars) erupt between about 6 and 13 years. Calcification of the permanent dentition begins in late gestation and continues to about 16 years, on average. Similar to the criteria for skeletal and sexual maturity, the stages of calcification are discrete and the criteria are somewhat arbitrary. The sex difference in dental maturation is less pronounced than for other maturity systems [38].

Correlations between dental (based on calcification, Demirjian method) and Tanner–Whitehouse I skeletal ages are generally low in children 7 to 13 years of age [36]. Dental maturity (the ages at which individuals attain 14, 20, and 26 permanent teeth) is generally independent of sexual, skeletal, and somatic maturity during male adolescence [39].

Interrelationships among maturity indicators

The issue of interrelationships among the various indicators of biological maturation is complex, because only skeletal maturity and percentage of adult stature span the entire maturation period from birth to adulthood. Indicators such as age at PHV, stages of sexual maturation, and age at menarche in girls are limited

to puberty. A cluster analysis of 21 maturity indicators (skeletal, sexual, somatic, and dental) assessed in a sample of 111 Polish boys followed longitudinally from 8 to 18 years identified a general maturity factor during adolescence. This general factor included ages at peak velocity for several linear dimensions, attainment of stages of sexual maturity, skeletal ages of 14 and 15 years, ages at attaining 90%, 95%, and 99% of adult stature, and age at onset of the growth spurt in height. Correlations among these indicators were high; none was below 0.70 and many were above 0.80. This suggests central regulation of the timing of the growth spurt and sexual maturation by the nervous system and corresponding hormonal correlates.

The second and third factors were related to indicators associated with prepubertal maturity (skeletal age of 11 and 12 years, 80% of adult height) and the ages by which 14, 20, and 26 teeth had erupted [39]. Similar results were obtained in Polish girls [40] and in American boys and girls [41], although indicators of dental maturity were not included in these analyses. The clustering of prepubertal maturational events that are somewhat independent of the clustering of pubertal events suggests that different hormonal and related growth factors are the driving forces that underlie these events. In general, it is the hypothalamic–pituitary–GH/IGF-I and the hypothalamic-pituitary-gonadal axes, but especially their interactions, that drive adolescent growth and maturation, given adequate thyroid status.

Indicators of skeletal, somatic, and sexual maturity are thus related during adolescence. When children are grouped according to an event of sexual maturation, the mean chronological age and the skeletal age at reaching that event are generally quite similar, but the standard deviation in skeletal age at reaching the event is markedly reduced. There is more variation in chronological age than in skeletal age at the time of menarche and at the time of PHV [3].

Timing, sequence, and tempo of maturational events

Overview

The mean and median ages at reaching various stages of somatic and sexual maturation are summarized in **tables 1** and **2**, respectively. The age at takeoff of the adolescent growth spurt averages

TABLE 1. Mean age (years) at takeoff and at peak height velocity (PHV) in samples of European and North American adolescents[a]

	Girls		Boys	
Population	Takeoff	PHV	Takeoff	PHV
Europe	8.2–10.3	11.4–12.2	10.3–12.1	13.8–14.4
North America				
Caucasian	8.7–9.6	11.3–12.0	10.5–11.4	13.3–14.1
African-American	8.9	10.8	10.3	14.3

a. Adapted from Malina et al. [3, 42] and Beunen and Malina [43].

TABLE 2. Median/mean ages at the onset of stages of sexual maturation in samples of European and North American adolescents[a]

	Girls' breast stage		Girls' pubic hair stage	
Population	B2	B5	PH2	PH5
Europe	10.0–11.6	14.0–15.7	10.4–12.1	13.6–15.4
North America				
Caucasian	10.0–11.2	13.7–15.5	10.5–11.6	13.1–16.3
African-American	8.9–9.5	13.9	8.8–9.5	14.7
Mexican-American	9.8–10.9	14.7	10.4–10.5	15.5–16.3
	Boys' genital stage		Boys' pubic hair stage	
Population	G2	G5	PH2	PH5
Europe	10.8–11.4	14.9–16.1	11.5–13.4	14.9–16.0
North America				
Caucasian	10.0–11.8	14.3–17.3	11.2–12.2	14.3–16.1
African-American	9.2	15.0	11.2	15.3
Mexican-American	10.3–12.4	15.8–16.3	12.3–16.3	15.7–16.1

a. Adapted from [3].

between 8.0 and 10.3 years in samples of European and North American girls, and the age at PHV is about 2 years later (10.8 to 12.2 years). Corresponding maturational events occur about 2 years later in boys. The standard deviations of the somatic maturity characteristics vary between 0.7 and 1.2 years, indicating a high degree of interindividual variation in the timing of the growth spurt. The mean age ranges of boys and girls from Europe and North America (Caucasian, African-American) overlap.

The mean or median ages at reaching breast stage 2 (B2) vary between 8.9 and 11.6 years and are earlier in African-Americans. Similar ethnic differences are apparent for breast stage B5 and pubic hair stages PH2 and PH5. B2 is, on average, the first overt sign of puberty in girls, and genital stage 2 (G2) is the first overt sign in boys. G2 occurs between the ages of 9.2 and 12.4 years and is also somewhat earlier in African-Americans. Note, however, that the appearance of pubic hair (PH2) may precede breast or genital development. The standard deviations of age at reaching stages of sexual maturation are generally larger than those for age at PHV and are larger for the more advanced stages. The latter may reflect difficulties in assessing stages 3 through 5 of breast, genital, and pubic hair development. The average age at menarche is between 12.1 and 13.5 years in European and North American girls. African-American girls attain menarche earlier than Caucasian girls, and within Europe there is a north–south gradient, with the mean age at menarche declining from north to south. Variation within and between countries is relatively large, with standard deviations of about 1 year. It should be noted that interindividual variation within populations is considerable.

The transition from one stage to the next is an indicator of the *tempo* of maturation. However, longitudinal data documenting the duration of stages are very limited. The duration of the pubertal transition from G2 to G5, B2 to B5, and PH2 to PH5 is quite variable. The average duration was about 2.2 years for breast development and 2.7 years for pubic hair development in Swiss girls from the Zurich Longitudinal Study. The corresponding estimates for Swiss boys are, on average, 3.5 years for genital development and 2.7 years for pubic hair development. The standard deviation is about 1.0 year [44, 45]. Data from the Harpenden Growth Study indicate longer durations, 4.0 years for breast and 2.5 years for pubic hair development. Note, however, that the 95th percentile for breast development from B2 to B5 is almost 9.0 years, whereas the 5th percentile is 11.5 years [18]. Some of the extreme variation in the Harpenden Growth Study may be due to methodological limitations of assessing the development of secondary sex characteristics from photographs. The setting of the Harpenden Growth Study was a children's home. Although the children were well cared for at the home, most of them had probably lived under socially disadvantageous conditions early in life. It is, however, difficult to assess the impact of these disadvantageous conditions early in life on the timing and sequence of adolescent events, especially at the individual level. Nevertheless, the broad range of variation in timing and tempo implies major limitations on the use of the average sequence of development of biological maturity indicators.

Factors that affect the timing, sequence, and tempo of maturational events

Although the processes of biological maturation and corresponding indicators are under strong genetic control (see related chapter by Thomis and Towne [46] in this issue), a number of environmental factors are also associated with variation in maturation. Chronic undernutrition is perhaps the most significant. It is often associated with impoverished social and economic conditions. Other factors include social class variation in some developed countries, familial characteristics, climate, altitude, and disease.

Undernutrition is associated with later ages at PHV and menarche in rural areas of developing countries [3]. Skeletal age is more delayed relative to chronological age in undernourished than in well-nourished children [12]. There is, however, variation among studies in the extent of delay in skeletal age, depending on the method of assessment. For example, Fels skeletal ages are significantly delayed relative to Tanner–Whitehouse II skeletal ages and relative to chronological age in school-aged Mexican children living under impoverished health and nutritional circumstances [47]. The results suggest that some of the variation in skeletal maturity status among chronically undernourished children may reflect variation in methods of assessment.

Variation in maturity status between ethnic or racial groups is less pronounced than that within populations, especially variation between the undernourished and well-nourished or between economic extremes. The mean age at menarche of well-nourished girls from Africa and Asia varies between 12.4 and 13.6 years, values similar to those observed in European and North American girls. However, the mean age at menarche of undernourished girls or girls living in rural areas in some developing countries varies between 13.9 and 14.6 years [48, 49]. In reports published after 1980, differences in the mean age at menarche between African girls living in rural areas or under poor nutritional conditions and those from urban or better-off areas vary between 0.6 and 1.1 years. These differences are still larger than the differences in age at menarche between most African countries [50]. Similar but less pronounced differences have been reported for ethnic variation in skeletal age and age at PHV, but the data are limited to North America, Europe, and Japan [3, 51,

52]. Given ethnic or racial variation in maturity status, it is essential that samples from diverse populations be included in the development of an international growth reference. Presumably, the use of samples from North America or Europe, South America, Africa, and Asia (Near, Middle, and Far East) would result in a good representation for an international reference. The Tanner–Whitehouse III method [13] appears to be a reasonable international reference for skeletal maturity; it is based on samples from Europe (Belgium, Italy, Spain, and the United Kingdom), Asia (Japan), Latin America (Argentina), and North America (well-off children from the northern suburbs of Houston, Texas, USA).

Overweight and obesity result from an imbalance between energy intake and energy expenditure. Regardless of etiology, obesity is, on average, associated with advanced maturation among children and adolescents. Some evidence suggests that maturational timing apparently has a greater long-term effect on the level of fatness than the level of fatness has on maturational timing [53].

Although it is well documented that elite female athletes in several sports are characterized by late biological maturation, there is no convincing evidence that systematic physical activity or regular training for sports has a causal influence on the timing of maturation [1–3]. Chronically low energy availability, which is sometimes observed in elite athletes, may contribute to later maturation, but this has not been established [54]. Nevertheless, chronically low energy availability is probably a causal factor in the regulation of reproductive function in mature adolescents and adult women [54].

In contrast to measures of body size, variation in the ages at PHV and menarche associated with socioeconomic status is generally smaller. Among Polish and British youths, those from better-off socioeconomic circumstances attain PHV and menarche somewhat earlier than those from poorer conditions. However, the ages at PHV and menarche do not consistently differ among Swedish adolescents grouped according to socioeconomic status [3]. Urban–rural contrasts in indicators of maturity status are apparent in several European countries (e.g., Poland and Greece); they are negligible in others [3]. Urban–rural differences in less-developed countries are larger and probably reflect socioeconomic status and nutritional factors [55]. The age at menarche is also related to family size, increasing by 0.1 to 0.2 years for each additional sibling in the family among both nonathletic and athletic European and North American girls [56].

The mean age at menarche has a moderate negative correlation (–0.5 to –0.6) with the mean annual temperature of the habitat [57]. On the other hand, the association between altitude of residence and maturation varies among racial or ethnic groups. Children living at high altitudes in the Andes mature later than those living in the lowlands, but the opposite is observed in Ethiopia [3]. These observations may be explained, in part, by variation in living conditions (e.g., nutritional conditions, infectious disease load, and poor public health) associated with lower socioeconomic status in the ethnic groups residing at high altitude.

Since many of the factors that can influence biological maturation process are interrelated, it is difficult to partition independent effects. Nevertheless, factors that potentially have an adverse effect on maturity status should be considered in the inclusion or exclusion criteria in the development of an International Growth Standard for Preadolescent and Adolescent Children. This can be perhaps be achieved by selection of adequate subsamples.

Seasonal variation in maturity

Since information on the topic of seasonal variation in maturity is very limited [58, 59], it will not be covered in this review. Among Canadian boys and girls 8.5 to 18.0 years of age, about 67% and 60% of the yearly growth in height, respectively, was accounted for by summer velocities [60].

Design of studies of maturational events

On the basis of experience in planning studies of growth and maturation, present knowledge about variation in the timing and sequence of maturational events, and methodological considerations in constructing reference data [61], the following recommendations are offered:

» Cross-sectional designs can be used to construct reference data for maturational events using the status quo method [5];
» Longitudinal data are required to obtain precise information about growth and maturation patterns [5];
» Longitudinal observations made every 3 months are optimal for describing maturation during adolescence. It can be verified whether observations made every 6 months provide accurate data on maturation. This may be done by using already available longitudinal data from observations made every 3 or 4 months;
» Longitudinal observations should be made from preadolescence onwards. Given interindividual variation, the observations should start at a fairly early age, most likely from 8 years onwards in girls and starting a year later in boys. Some data indicate that a significant percentage of US girls may begin puberty at even earlier ages [62], which suggests that it may be advisable to start even at 6 years in girls and a year or so later in boys. If it is feasible, ultrasensitive estrogen assays based on molecular biological techniques can be used to accurately predict the onset of

pubertal development before the external signs (e.g., breast development) appear. This could considerably reduce the length of the follow-up needed to cover the adolescent maturation period;

» A pure longitudinal study over a relatively long period with four measurement periods per year may not be feasible; a multiple (mixed) longitudinal design may provide accurate results. Such a design could consist of a follow-up of birth cohorts 1 year apart that are followed over 1 year (with two or four measurements per year) or, more likely, several cohorts followed over 3 to 4 years with overlapping age levels (at least one age level). Again, the efficiency of such a mixed longitudinal design should be verified by data from pure longitudinal studies. The results of the simulated mixed longitudinal design should then be validated against the pure longitudinal data;

» The sample size depends on the variable, age, and percentiles required and whether the distribution is normal or can be normalized. For body-mass index (BMI), Guo* demonstrated that the confidence limits of 95th percentiles markedly decrease until the sample size reaches 200 subjects; see also the article by Cole [63] in this volume.

» A combination of cross-sectional and longitudinal designs should be considered.

Secular change in maturational events

Summary of secular changes until 1970

Most of the available evidence for secular changes in biological maturation is derived from records of the age at menarche. Data from retrospective and status quo techniques do not necessarily correspond closely [49]. Although most of the more recent publications are based on the status quo method, older data are partly or entirely based on retrospective data (for a more detailed discussion see Danker-Hopfe [64]).

The mean age at menarche in Norwegian girls was rather stable at 16 years in lower social strata and 14 years in higher social strata from 1820 to 1910–20 and declined subsequently to 13.3 years in the early 1950s [65, 66]. In the United States, the mean age at menarche declined from about 14.7 years in the 1870s to 12.8 in the 1950s [67]. Corresponding data for Japan indicate a decline from a bit over 16.0 years at the end of the nineteenth century to about 15.0 years in girls born around 1930 and subsequently to 13.0 years for girls born after World War II [3].

Recent secular trends in age at menarche

The trend toward earlier menarche has slowed or stopped in several countries. Since the 1960s, changes have been small in US girls, about 0.2 years in European Americans and about 0.4 years in African-Americans. The trend has also stopped or slowed in several European countries, such as the United Kingdom, Netherlands, Hungary, the former German Democratic Republic, Croatia, and Portugal [64, 68]. Recent reports demonstrate that the positive secular decline continues in Denmark [69] and South Korea [70]. Data from Poland illustrate a social gradient and secular change. The mean age at menarche declined somewhat more in girls living in urban conditions than in girls living in towns and villages. Between 1966 and 1978, however, the secular decline was more marked in girls from towns and villages than in urban girls. Subsequently, the mean age at menarche increased from 1978 to 1988; the increase was greatest in girls from towns and least in girls from villages. The recent negative secular trend was probably related to political, social, and economic conditions [71, 72].

Data from longitudinal studies in Europe spanning 50 years and from the United States spanning 75 years provide estimates of age at PHV. In Europe the age at PHV varied between 13.8 and 14.2 years for boys in 25 of 26 samples and between 11.6 and 12.3 years for girls in 24 of 25 samples. In the United States, the age at PHV varied between 11.3 and 11.9 years in girls and between 13.3 and 14.1 years in boys. With allowance for differences in the method of estimating age at PHV, sampling errors, and the uniqueness of the longitudinal samples, these data suggest no clear secular trend [3].

In contrast, data from the annual School Health Surveys conducted in Japan show a gradual decline in maximal increment age (MIA), which is similar but not identical to age at PHV. MIA declined (positive secular change) from the beginning of the twentieth century and subsequently increased (negative secular trend) during World War II and the years immediately thereafter; subsequently, the decline continued through the 1990s. Overall, the estimated rate of the trend has slowed between 1960 and 1990 [73]. Similar changes were observed in Taiwan and in mainland China [3].

Factors that affect secular changes

Many reasons have been postulated for the trend toward earlier maturity, but the underlying causes are not known with certainty. It is reasonable to assume that many interrelated factors are involved, especially the elimination of growth-inhibiting factors. Improved living conditions, sanitation, and overall public health, as reflected in the marked reduction in infant and childhood mortality and morbidity, are primary contributors [49, 64, 74]. Improved nutrition and associated beneficial changes in public health are related factors [3]. Although genetic changes have also been postulated, secular changes occur too rapidly to be accounted for by genetic changes in a population [49, 74]. Decline in family size, increased sexual

* National Center for Health Statistics. Executive summary of the growth chart workshop 1992. Hyattsville, Md, USA: Centers for Disease Control and Prevention, 1994.

stimulation, and decreased "pastoralization" (raising livestock as a primary economic activity) have also been suggested as contributing factors [3, 64].

Recommendations for the construction of an international growth standard for preadolescent and adolescent children

Because biological maturation is closely related to growth, it is of relevance in monitoring the growth status of children and adolescents and also in screening of children at risk. Thus, it is important to include indicators of biological maturation in all growth studies [3, 5, 75].

With allowance for the limitations and advantages of the different indicators of biological maturation considered in this chapter, it is recommended that the following be included: indicators of sexual maturation (stages of pubic hair and breast development and age at menarche in girls, and stages of pubic hair and genital development in boys); indicators of the adolescent growth spurt (age at takeoff and age at PHV); and skeletal maturity. If possible, it would be helpful to have samples of saliva or blood to measure the stable levels of steroid hormones and perhaps IGF-I. Measures of the pulsatile nature of the peptide hormones are entirely impractical and would probably show more variability than the physical measures. Information concerning the state of the hypothalamic–pituitary–gonadal axis, including ovarian cycles, can be obtained from the concentrations of pregnanediol glucuronide, estrone conjugates, and the gonadotropins measured in urine by specific chemiluminescent assays [76].

A cross-sectional design is adequate for the construction of reference data. However, if an accurate description of growth and maturation patterns is desirable, a longitudinal or mixed longitudinal design is required.

In closing, it should be noted that among auxologists opposite views have been expressed with regard to the construction and practicality of a universal growth reference:

> This diversity is important if genetic differences cause growth variations, but a considerable literature indicates differences cause only a minor part of the growth variances between populations. This implies that a single set of reference data could be used internationally if it were obtained by excellent procedures from a *population free of retarding influences* [emphasis added] (Roche [75], p. 80)
>
> In regard to standards for individuals, it used to be said that the growth of all healthy populations, at least up to age five was about the same and one universal standard would do for all. The data in this book make it plain that this is a misconception, based on an inadequate sample of populations…. Clearly what is needed—and what is very actively in progress—is for countries, or at least broad regions, to generate their own standards. These should be based on *well-nourished healthy individuals*, [emphasis added] or the nearest approach to that ideal that is practicable, and if used over adolescence they should be longitudinal and have separate channels for early and late maturers (Eveleth and Tanner [49], p. 15).

The material presented in this volume must provide a sound basis to decide which of these positions is based on sound evidence presently available.

References

1. Beunen G. Physical growth, maturation and performance. In: Eston R, Reilly T, eds. Kinanthropometry and exercise physiology laboratory manual. Vol 1: Anthropometry. Tests, procedures and data. London: Routledge, 2001:65–90.
2. Beunen G, Malina RM. Growth and biological maturation: relevance to athletic performance. In: Bar-Or O, ed. The child and adolescent athlete. Encyclopaedia of sports medicine, vol VI. Oxford, UK: Blackwell, 1996:3–24.
3. Malina RM, Bouchard C, Bar-Or O. Growth, maturation, and physical activity, 2nd ed. Champaign, Ill, USA: Human Kinetics, 2004.
4. Roche AF. Bone growth and maturation. In: Falkner F, Tanner JM, eds. Human growth. Vol 2. Postnatal growth, neurobiology. New York: Plenum, 1986:25–60.
5. Tanner JM. Growth at adolescence, 2nd ed. Oxford, UK: Blackwell, 1962.
6. Acheson RM. Maturation of the skeleton. In: Falkner F, ed. Human development. Philadelphia, Pa, USA: Saunders, 1966:465–502.
7. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist, 2nd ed. Palo Alto, Calif, USA: Stanford University Press, 1959.
8. Todd TW. Atlas of skeletal maturation. St Louis, Mo, USA: Mosby, 1937.
9. Roche AF, Wainer H, Thissen D. Skeletal maturity: the knee joint as a biological indicator. New York: Plenum Press, 1975.
10. Tanner JM, Whitehouse RH, Healy MJR. A new system for estimating skeletal maturity from the hand and wrist, with standards derived from a study of 2,600 healthy British children. Paris: International Children's Centre, 1962.
11. Tanner JM, Whitehouse RH, Marshall WA, Healy MJR, Goldstein H. Assessment of skeletal maturity and prediction of adult height (TW2 method). New York: Academic Press, 1975.
12. Tanner JM, Whitehouse RH, Cameron N, Marshall WA, Healy MJR, Goldstein H. Assessment of skeletal maturity and prediction of adult height. 2nd ed. New York: Academic Press, 1983.

13. Tanner JM, Healy MJR, Goldstein H, Cameron N. Assessment of skeletal maturity and prediction of adult height (TW3 method), 3rd ed. London: Saunders, 2001.
14. Roche AF, Chumlea WC, Thissen D. Assessing the skeletal maturity of the hand-wrist: Fels Method. Springfield, Ill, USA: Charles C Thomas, 1988.
15. Tanner JM, Gibbons RD. A computerized image analysis system for estimating Tanner-Whitehouse 2 bone age. Horm Res 1994;42:282–87.
16. Tanner JM, Oshman D, Lindgren G, Grunbaum JA, Elsouki R, Labarthe D. Reliability and validity of computer-assisted estimates of Tanner-Whitehouse skeletal maturity (CASAS): comparison with the manual method. Horm Res 1994;42:288–94.
17. Roemmich JN, Blizzard RM, Peddada SD, Malina RM, Roche AF, Tanner JM, Rogol AD. Longitudinal assessment of hormonal and physical alterations during normal puberty in boys. IV: Prediction of adult height by the Bailey-Pinneau, Roche-Wainer-Thissen, and Tanner-Whitehouse methods compared. Am J Hum Biol 1997;9:371–80.
18. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in girls. Arch Dis Child 1969;44: 291–303.
19. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child 1970;45:13–23.
20. Prader A. Testicular size: assessment and clinical importance. Triangle 1966;7:240–3.
21. Zachmann M, Prader A, Kind HP, Hafliger H, Budliger H. Testicular volume during adolescence. Cross-sectional and longitudinal studies. Helv Paediatr Acta 1974;29:61–72.
22. Fisher DA. Thyroid function in the fetus. In: Fisher DA, Burrow GN, eds. Perinatal thyroid physiology and disease. New York, Raven Press, 1975:21.
23. Underwood LE, D'Ercole AJ. Insulin and insulin-like growth factors/somatomedins in fetal and neonatal development. Clin Endocrinol Metab 1984;13:69–81.
24. MacLaughlin DT, Teixeira J, Donahoe PK. Perspective: reproductive tract development—new discoveries and future directions. Endocrinology 2001;142:2167–72.
25. Winter JS, Faiman C. Pituitary-gonadal relations in male children and adolescents. Pediatr Res 1972;6:126–35.
26. Griffin JE, Wilson JD. Disorders of the testes and the male reproductive tract. In: Larsen PR, Kronenberg HM, Melmed S, Polonsky KS, eds. Williams textbook of endocrinology, 10th ed. Philadelphia, Pa, USA: Saunders, 2003:709–69.
27. Zhang J, Peddada SD, Malina RM, Rogol AD. Longitudinal assessment of hormonal and physical alterations during normal puberty in boys. VI. Modeling of growth velocity, mean growth hormone (GH mean), and serum testosterone (T) concentrations. Am J Hum Biol 2000; 12:814–24.
28. Gasser T, Kohler W, Muller HG, Kneip A, Largo R, Molinari L, Prader A. Velocity and acceleration of height growth using kernel estimation. Ann Hum Biol 1984;11:397–411.
29. Hauspie R, Chrzastek-Spruch H. Growth models: possibilities and limitations. In: Johnston FE, Zemel B, Eveleth PB, eds. Human growth in context. London: Smith-Gordon, 1999:15–24.
30. Marubini E. Mathematical handling of long-term longitudinal data. In: Falkner F, Tanner JM, eds. Human growth. Vol 1. Principles and prenatal growth. New York: Plenum, 1978:209–25.
31. Bayley N, Pinneau SR. Tables for predicting adult height from skeletal age: revised for use with the Greulich-Pyle hand standards. J Pediatr 1952;40:423–41.
32. Roche AF, Wainer H, Thissen D. Predicting adult stature for individuals. Monogr Paediatr 1975;3:1–114.
33. Roche AF, Wainer H, Thissen D. The RWT method for the prediction of adult stature. Pediatrics 1975;56: 1027–33.
34. Khamis HJ, Roche AF. Predicting adult stature without using skeletal age: the Khamis-Roche method. Pediatrics 1994;94(4 Pt 1):504–7. Erratum in: Pediatrics 1995; 95:457.
35. Beunen GP, Malina RM, Lefevre J, Claessens AL, Renson R, Simons J. Prediction of adult stature and noninvasive assessment of biological maturation. Med Sci Sports Exerc 1997;29:225–30.
36. Demirjian A. Dentition. In: Falkner F, Tanner JM, eds. Human growth. Vol 2. Postnatal growth, neurobiology. New York: Plenum, 1986:269–98.
37. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. Hum Biol 1973;45:211–27.
38. Demirjian A, Buschang PH, Tanguay R, Patterson DK. Interrelationships among measures of somatic, skeletal, dental, and sexual maturity. Am J Orthod 1985;88:433–8.
39. Bielicki T, Koniarek J, Malina RM. Interrelationships among certain measures of growth and maturation rate in boys during adolescence. Ann Hum Biol 1984; 11:201–10.
40. Bielicki T. Interrelationships between various measures of maturation rate in girls during adolescence. Studies in Physical Anthropology 1975;1:51–64.
41. Nicolson AB, Hanley C. Indices of physiological maturity: derivation and interrelationships. Child Dev 1953;24:3–38.
42. Malina RM, Bouchard C, Beunen G. Human growth: selected aspects of current research on well nourished children. Annu Rev Anthropol 1988;17:187–219.
43. Beunen G, Malina RM. Growth and physical performance relative to the timing of the adolescent spurt. Exerc Sport Sci Rev 1988;16:503–40.
44. Largo RH, Prader A. Pubertal development in Swiss boys. Helv Paediatr Acta 1983;38:211–28.
45. Largo RH, Prader A. Pubertal development in Swiss girls. Helv Paediatr Acta 1983;38:229–43.
46. Thomis MA, Towne B. Genetic determinants of prepubertal and pubertal growth and development. Food Nutr Bull 2006;27(suppl):S257–78.
47. Pena Reyes RM, Malina RM. Fels and Tanner-Whitehouse skeletal ages of school children 7–13 years in Oaxaca, Mexico. In: Dasgupta P, Hauspie R eds. Perspectives in human growth, development and maturation. Dordecht, Netherlands: Kluwer Academic Publishers, 2001:55–65.
48. Cameron N, Wright CA. The start of breast development and age at menarche in South African black females. S Afr Med J 1990;78:536–9.
49. Eveleth PB, Tanner JM. Worldwide variation in human growth. 2nd ed. Cambridge, UK: Cambridge University Press, 1990.
50. Gillett-Netting R, Meloy M, Campbell BC. Catch-up

reproductive maturation in rural Tonga girls, Zambia? Am J Hum Biol 2004;16:658–69.
51. Kimura K. Studies on growth and development in Japan. Yearb Phys Anthropol 1984;27:179–214.
52. Murata M. Population-specific reference values for bone age. Acta Paediatr Suppl 1987;423:113–4.
53. Loucks AB. Energy availability, not body fatness, regulates reproductive function in women. Exerc Sport Sci Rev 2003;31:144–8.
54. Garn SM, LaVelle M, Rosenberg KR, Hauthorne VM. Maturational timing as a factor in female fatness and obesity. Am J Clin Nutr 1986;43:879–83.
55. Padez C. Social background and age at menarche in Portuguese university students: a note on the secular changes in Portugal. Am J Hum Biol 2003;15:415–27.
56. Malina RM, Katzmarzyk PT, Bonci CM, Ryan RC, Wellens RE. Family size and age at menarche in athletes. Med Sci Sports Exerc 1997;29:99–106.
57. Roberts DF. Climate and human variability. Module in anthropology no. 34. Reading, Mass, USA: Addison-Wesley, 1973.
58. Marshall WA. Evaluation of growth rate in height over periods of less than one year. Arch Dis Child 1971;46:414–20.
59. Reynolds EL, Sontag LW. Seasonal variation in weight, height and appearance of ossification centers. J Pediatr 1944;24:524–35.
60. Mirwald RL, Bailey DA. Seasonal height velocity variation in boys and girls 8–18 years. Am J Hum Biol 1997; 9:709–15.
61. Goldstein H. The design and analysis of longitudinal studies. London: Academic Press, 1979.
62. Herman-Giddens ME, Slora EJ, Wasserman RC, Bourdony CJ, Bhapkar MV, Koch GG, Hasemeier CM. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings Network. Pediatrics 1997;99:505–12.
63. Cole TJ. The International Growth Standard for Preadolescent and Adolescent Children: Statistical considerations. Food Nutr Bull 2006;27(suppl):S237-243.
64. Danker-Hopfe H. Menarcheal age in Europe. Yearb Phys Anthropol 1986;29:81–112.
65. Brundtland GH, Walloe L. Menarcheal age in Norway in the 19th century: a re-evaluation of the historical sources. Ann Hum Biol 1976;3:363–74.
66. Liestol K, Rosenberg M. Height, weight and menarcheal age of schoolgirls in Oslo—an update. Ann Hum Biol 1995;22:199–205.
67. Wyshak G, Frisch RE. Evidence for a secular trend in age of menarche. N Engl J Med 1982;306:1033–5.
68. Bodzsar EB, Susanne C, eds. Secular growth changes in Europe. Budapest, Hungary: Eotvos Lorand University Press, 1998.
69. Olesen AW, Jeune B, Boldsen JL. A continuous decline in menarcheal age in Denmark. Ann Hum Biol 2000; 27:377–86.
70. Hwang J-Y, Shin C, Frongillo EA, Shin KR, Jo I. Secular trend in age at menarche for South Korean women born between 1920 and 1986: the Ansan study. Ann Hum Biol 2003;30:434–42.
71. Bielicki T. Secular trends in growth: human biologists' contribution to the study of social change. In: Johnston FE, Zemel B, Eveleth PB, eds. Human growth in context. London: Smith-Gordon, 1999:303–11.
72. Bielicki T, Hulanicka B. Secular trend in stature and age at menarche in Poland. In: Bodzsar EB, Susanne C, eds. Secular growth changes in Europe. Budapest, Hungary: Eotvos Lorand University Press, 1998:263–79.
73. Ali MA, Ohtsuki F. Estimation of maximum increment age in height and weight during adolescence and the effect of World War II. Am J Hum Biol 2000;12:363–70.
74. Malina RM. Secular changes in size and maturity: causes and effects. Monogr Soc Res Child Dev 1979;44(179):59–102.
75. Roche AF. Growth, maturation and body composition. The Fels Longitudinal Study 1929–1991. Cambridge, UK: Cambridge University Press, 1992.
76. Santoro N, Crawford SL, Allsworth JE, Gold EB, Greendale GA, Korenman S, Lasley BL, McConnell D, McGaffigan P, Midgely R, Schocken M, Sowers M, Weiss G. Assessing menstrual cycles with urinary hormone assays. Am J Physiol Endocrinol Metab 2003;284: E521–30.

Genetic determinants of prepubertal and pubertal growth and development

Martine A. Thomis and Bradford Towne

Abstract

This article surveys the current general understanding of genetic influences on within- and between-population variation in growth and development in the context of establishing an International Growth Standard for Preadolescent and Adolescent Children. Traditional genetic epidemiologic analysis methods are reviewed, and evidence from family studies for genetic effects on different measures of growth and development is then presented. Findings from linkage and association studies seeking to identify specific genomic locations and allelic variants of genes influencing variation in growth and maturation are then summarized. Special mention is made of the need to study the interactions between genes and environments. At present, specific genes and polymorphisms contributing to variation in growth and maturation are only beginning to be identified. Larger genetic epidemiologic studies are needed in different parts of the world to better explore population differences in gene frequencies and gene–environment interactions. As advances continue to be made in molecular and statistical genetic methods, the genetic architecture of complex processes, including those of growth and development, will become better elucidated. For now, it can only be concluded that although the fundamental genetic underpinnings of the growth and development of children worldwide are likely to be essentially the same, there are also likely to be differences between populations in the frequencies of allelic gene variants that influence growth and maturation and in the nature of gene–environment interactions. This does not necessarily preclude an international growth reference, but it does have important implications for the form that such a reference might ultimately take.

Key words: Association studies, heritability, linkage studies, population variation

Martine A. Thomis is affiliated with the Research Center for Exercise and Health, Department of Biomedical Kinesiology, Faculty of Kinesiology and Rehabilitation Sciences, Katholieke Universiteit Leuven, Leuven, Belgium; Bradford Towne is affiliated with Lifespan Health Research Center, Department of Community Health, Wright State University School of Medicine, Kettering, Ohio, USA.

Please address queries to the corresponding author: Martine A. Thomis, Research Center for Exercise and Health, Department of Biomedical Kinesiology, Faculty of Kinesiology and Rehabilitation Sciences, Katholieke Universiteit, Leuven, Tervuursevest 101, B-3001 Leuven, Belgium; e-mail: martine.thomis@faber.kuleuven.be.

Introduction

The aim of this paper is to provide a brief overview of traditional relative-based genetic epidemiologic analysis methods and to describe the findings from a sampling of studies that have quantified in some manner the contribution of genetic influences to variation in common measures of normal growth and development. This paper does not include discussion of genetically determined clinical growth disorders. Sources of systematic variation in human growth (e.g., secular trends and socioeconomic influences on growth) is discussed by Ulijaszek [1] in this supplement. Specifically, this paper focuses on current general understanding of genetic influences on within- and between-population variation in growth and development so that this knowledge can be considered in the broader discussion of the feasibility of establishing an international growth reference.

Methodological considerations

Measures of growth and development

Commonly collected measures of the growth and development of prepubertal and pubertal children consist of growth in height and weight, and indicators of morphological, skeletal and sexual maturity. A short description of these measures is given in **table 1**. In this supplement, Beunen et al. [2] discuss in detail

methodological issues pertaining to the assessment of morphological, skeletal, and sexual maturity characteristics during prepubertal and pubertal growth.

Methods for assessing genetic influences on growth and development

Familial correlations and heritability estimation

Two major study designs are typically used to initially quantify genetic and environmental influences on measures of growth and development: twin studies and family studies. In most modern twin studies, both monozygotic (MZ) and dizygotic (DZ) twins are examined. Since MZ twins have an identical genetic makeup, they will be more similar (i.e., exhibit a higher intrapair correlation) in a trait (that is under at least some genetic control) than DZ twins, who share on average half of their genes (as do full siblings from singleton births). When twin data are used, genetic influences, as well as environmental influences both unique to the individual and shared within families, can be identified, as can, when certain assumptions are made, dominant genetic effects.

In family studies, patterns of similarities in a trait among parents and offspring, siblings, and other pairings of relatives are evaluated. This approach allows for the quantification of genetic and nongenetic sources of variation in a trait (including specifically identified nongenetic sources of variation). If data from combined pedigrees (e.g., twins and their parents) or extended families (e.g., grandparents, children, and grandchildren) are available, more sophisticated models can be tested, because such data are from relatives of varying degrees of relationship to each other and who span multiple generations and, oftentimes, households. Detailed discussion of analysis of twin and family data be found in basic texts (e.g., Neale and Cardon [9]).

Regardless of study design, the resulting heritability (h^2) of a trait is the key first step toward understanding the nature of genetic influences on a trait. Detailed discussion of quantitative genetic analysis can be found in basic textbooks [10, 11]. Briefly, the heritability of a trait is a measure of the extent to which the variation observed in a trait can be ascribed to genetic factors, ranging from 0% (no genetic effects) to 100% (complete genetic control). Assessment of heritability is based initially on a simple quantitative genetic model in which the total variation (V_P) in a trait is partitioned into genetic (V_G) and environmental (V_E) variance components such that $V_P = V_G + V_E$. Broad-sense heritability specifically refers to the proportion of the total phenotypic variation that can be attributed to all genetic effects (primarily consisting of additive, dominance, and epistatic effects) as defined by V_G/V_P (i.e., $h^2 = V_G/V_P$). The variance term V_G can be decomposed into an additive genetic term (V_A) that refers to the additive effects of genes across several genetic loci, a dominance

TABLE 1. Common measures of prepubertal and pubertal growth and development

Characteristic measured	Variable	Assessment
Somatic growth	Height (age 0 yr–adult)	Measured/self-reported (cm)
	Weight (age 0 yr–adult)	Measured/self-reported (kg)
Morphological maturation	Timing Age at take-off of growth spurt (i.e., prepubertal nadir)	Derived from curve-fitting approaches (Preece–Baines I [3], triple logistic model [4, 5]) (yr)
	Age at peak height velocity during puberty	Derived from curve-fitting approaches (Preece–Baines I, triple logistic model) (yr)
	Tempo Height velocity at age at take off of growth spurt	Derived from curve-fitting approaches (Preece–Baines I, triple logistic model) (cm/yr)
	Height velocity at peak height velocity	Derived from curve-fitting approaches (Preece–Baines I, triple logistic model) (cm/yr)
Skeletal maturation	Skeletal age at chronological age 1–18 yr	Tanner–Whitehouse II method [6] (yr) or Fels method [7] (yr)
	Skeletal maturation score at different chronological ages	Tanner-Whitehouse II method (score in points)
Sexual maturation	Age at menarche	Retrospective self-report or prospective self-report (yr)
	Breast development	Breast development stages B1-B5 (Tanner [8]), clinical assessment or self-report
	Pubic hair development	Pubic hair development stages PH1–PH6 (Tanner [8]), clinical assessment or self-report
	Genital development	Genital development stages G1–G5 (Tanner [8]), clinical assessment or self-report

term (V_D) that refers to the interaction between alleles at the same locus (i.e., the heterozygote effect is not intermediate between the two homozygote genotype effects), and an epistasis term (V_I) that refers to the interaction between alleles at different loci. Narrow-sense heritability specifically refers to the proportion of the total phenotypic variation that can be attributed only to the additive effects of one or more genes at one or more chromosomal locations (i.e., V_A) and is defined as $h^2 = V_A/V_P$. In most instances, heritability estimates refer to narrow-sense heritabilities. Generally speaking, narrow-sense heritability is of more utility in characterizing genetic influences on continuously distributed traits such as various measures of growth and development. Variation in such traits is probably influenced by a number of genes, each with small to moderate effects.

In this variance components framework, the environmental variance term V_E can similarly be decomposed into components due to, for example, environmental factors shared by certain family members (e.g., common household environment or specific aspects thereof) and/or components due to environmental factors specific to the individual (e.g., behaviors or habits unique to certain family members).

The quantitative genetic models of sources of variation outlined above make several assumptions that may or may not be realistic, depending upon the particular study population and traits being examined. These assumptions include no gene–environment interaction (i.e., different genotypes all react equally to environmental factors), no gene–environment correlation (i.e., different genotypes are distributed equally across different environments), no gene–gene interaction (i.e., a genotype at one locus that contributes to variation in a trait has no influence on the effect that a genotype at another locus might have on variation in that same trait), and no assortative mating (i.e., people mate randomly with respect to the phenotype in question). Violation of these assumptions does not invalidate the fundamental approach (such violations may or may not be significant in magnitude) but requires that caution be used in interpreting results. It must also be kept in mind that heritabilities are population-level estimates, specific to a particular population in a particular environment, and this can sometimes be an important consideration when heritability estimates across populations are compared. For example, differences in heritability estimates in groups from different ethnic backgrounds, of different socioeconomic status, or from affluent versus developing countries can be indicative of differences in the relative importance of genetic and environmental factors in different circumstances [12, 13]. For example, Luke et al. [14] found significant differences among Nigerian, Jamaican, and US blacks in the heritability of adult height (h^2 = .62, .74, .87, respectively), although they did not speculate as to the reasons for these differences. In that same study, obesity-related phenotypes showed large differences in the prevalence of obesity (5.4% in Nigerians, 23.3% in Jamaicans, and 39.0% in US blacks), but their heritability estimates did not differ among the three groups. In general, heritability estimates for various traits, including measures of growth and development, tend to be robust and generally similar across populations.

Incorporation of environmental factors into genetic analyses

Because of significant differences between populations in lifestyle (e.g., diet and physical activity), environmental exposures (e.g., disease), and evolutionary histories (e.g., different selective pressures), gene–environment interactions merit special attention in the study of genetic influences on growth and development. Gene–environment interaction is probably an important influence on the variation observed among children in their growth and development, particularly in populations exposed to environmental factors known to negatively impact health. The key to gene–environment interaction is that not all children may respond to the same degree to such environmental factors, and a portion of that differential response may be due to genetic variation among them.

The study of gene–environment interaction effects on complex traits using modern molecular and statistical genetic methods, however, is still in its infancy, and is likely to constitute a major component of genetic epidemiologic research for decades to come. To study gene–environment interaction effects on growth and development requires data from a large number of related children who are exposed to an environmental factor shown to, on average, negatively (or positively) impact health. Although the principles of gene–environment interaction have been empirically demonstrated in a number of experimental organisms, and in a few instances in studies of humans using identical twin study designs, study of gene–environment interaction effects on the growth and development of children usually requires taking advantage of particular circumstances.

For example, in the Jiri Growth Study [15, 16], a cohort of more than 1,000 related children from the Jirel ethnic group in rural eastern Nepal, where intestinal helminthic infections are endemic, are being examined annually. Data collected include an extensive battery of anthropometrics, skeletal age assessments, and quantification of parasite burden. The central hypothesis of the Jiri Growth Study is that susceptibility to helminthic infections is genetically mediated and that this, in turn, contributes to patterns of variation observed among children in different measures of their growth and development. Similarly, the rapid increase in childhood overweight and obesity, particularly in developed nations, offers researchers opportunities for the study of gene–environment interactions. The vast majority of children in developed nations live in environments

where food is plentiful and relatively inexpensive. For reasons that have yet to be fully explored, however, not all children partake equally in this abundance. And, even among children who equally overindulge in this modern bounty, it is reasonable to hypothesize that there are differential responses in terms of weight gain and other measures of health status.

Linkage and association studies

As it has become well established that essentially all commonly collected measures of growth and development have significant and substantial genetic underpinnings, attention in the postgenomic era has turned to the identification of specific genes that influence variation in those measures. Linkage analysis and association studies are two complementary strategies [17] used to identify genes that explain a portion of the variability in complex traits, including measures of growth and development.

Linkage analysis is an important initial tool for the mapping of genetic loci harboring genes that influence the trait of interest. In a modern whole genome linkage scan, related individuals are genotyped for several hundred highly variant DNA markers whose exact chromosomal locations are known and that are regularly spaced throughout the genome (e.g., every 10 centimorgans). Different approaches to linkage analysis exist [17], but the fundamental premise of any linkage analysis is that two loci physically close to each other on a chromosome are likely to be inherited together. As the distance between loci increases, the likelihood that these loci will cross over or recombine during meiosis (and thus not be inherited together) also increases. Loci close to each other on a chromosome are referred to as being linked. Loci harboring genes that influence variation in quantitative traits such as stature, weight, and the timing and tempo of growth are called quantitative trait loci (QTL). Linkage between a QTL and a marker exists when pairs of relatives who are more phenotypically similar share more alleles at a particular marker locus than pairs of relatives who are more phenotypically dissimilar.

The second approach to gene identification is association studies [17]. In association studies, the effect of a specific (polymorphic) marker allele (typically within a candidate gene) on variation in a trait is measured. For example, a study sample is divided into two or more groups on the basis of marker genotype, and differences between the groups in the trait being examined are evaluated by statistical tests such as analysis of variance (ANOVA). Testing for significant differences in marker allele frequencies in a case–control design can also be conducted. In this case, the test is for whether or not there are significant differences in marker allele frequencies between an "affected" group and a "control" group. As in linkage studies, in association studies the markers used may be single tandem repeat markers (STRs) or single nucleotide polymorphisms (SNPs) and can also include insertion/deletion polymorphisms or multiallelic variants. When a significant association is found (e.g., more girls with an early age at menarche being found to have a particular marker genotype than random expectation), the putative "early menarche" allele under study might be a true functional variant (e.g., resulting in an amino acid change in protein structure) or might be in linkage disequilibrium with the true functional allele (i.e., tending to be inherited together with the functional allele). As knowledge of the human genome has progressed, multiple polymorphisms within one or more genes, or even genome-wide, can now be studied. Here, instead of testing for association with each polymorphism separately, haplotypes (e.g., a set of SNPs along a region of a chromosome) can be tested for patterns of association with variation in the trait of interest.

Linkage and association studies are best viewed as being complementary to each other, or as different steps in the systematic intellectual endeavor of proceeding from observing variation in a trait to identifying the functional genetic determinants contributing to that variation [15, 18, 19]. Association studies have appeal because data from unrelated individuals can be used (however, family data can be included in the analyses to overcome problems of hidden population stratification, which can increase the finding of false positive associations). Unfortunately, the success of candidate gene association studies in identifying specific polymorphisms influencing complex traits has been decidedly mixed. Current genome-wide association studies hold promise, using SNP (single nucleotide polymorphism) microarray chips with up to 500,000 variants to densely cover the genome, but present their own set of problems (e.g., multiple testing). Linkage studies require data from (ideally) a large number of relatives to provide adequate statistical power for establishing linkage, and the cost of genotyping each family member for approximately 400 genetic markers (for the typical 10-centimorgan map) can be high (although the cost of genotyping has markedly decreased over the last decade). Once such a familial study population is established, however, a wide variety of traits can be examined using that same resource. A current common strategy for genetic epidemiologic studies of normal variation in complex traits proceeds from heritability estimation, to linkage analysis, to fine mapping, to sequencing, to final identification of a functional genetic polymorphism. Association studies in larger samples can then proceed in an informed manner to evaluate the effects of specific genes on the trait of interest in different populations. At each step in this modern process, the effects of covariates and of various types of interactions can be incorporated into the analyses. A shortcut to this stepwise procedure is to find specific candidate genes for growth-related

phenotypes. These can come from whole-genome gene expression studies (mRNA levels in tissue), although this might be unrealistic to perform because of the nature of the sampling procedures needed to obtain growing tissue material (e.g., bone, growth plate, muscle, and fat) in children and adolescents. Differences in up- or down-regulation of genes in growing tissue versus adult tissue might also be informative for the specific genes involved in regulating individual differences in growth. Candidate genes might also be detected from gene effects on growth and development in rodent or other animal models that might have similar effects in growing children.

Evidence for genetic effects on growth and development

Familial correlations and heritabilities

The literature on familial or genetic contributions to variation in measures of normal growth and maturation is fairly extensive and dates back almost a century. On the basis of that body of work, it is well established that genes play a major, if not the major, role in explaining variation observed among children in their growth and development, even in poor environments.

Stature, weight, and patterns of growth

Stature. Estimates of the contribution of genetic factors to interindividual differences in adult height have recently been reviewed [12, 20]. Detailed reviews of genetic influences on the growth and development of children have been published by Maes [21] and more recently by Towne et al. [15].

In general, in populations around the world, h^2 estimates of height range from 0.6 to more than 0.9, clearly showing that height is a highly heritable trait. Intrapair correlations for the recumbent lengths of MZ and DZ twins at birth are similar, which also suggests the importance of shared environmental factors (i.e., maternal and intrauterine factors) in contributing to the length of newborns. The similarities between MZ twins increase from 3 months until they reach a stable and high level of similarity at 3 years of age and onward. In DZ twins, the high similarity observed at birth decreases to a value approximately half that of MZ twins by the age of 3 years. Data on adult height from large twin cohorts of whites in eight developed countries show that although there are differences in population averages in adult height, in general only minor differences in the genetic architecture of height are present [20]. Although heritability estimates might be comparable in affluent populations, Mueller et al. [22] compared 24 studies of parent–child correlations and found that population estimates of heritability tend to be systematically lower in developing countries than in developed countries. Part of this observation might be explained by higher nutritional and disease stresses in non-European populations in the developing world. Such environmental factors have the potential to negatively affect growth, thereby increasing the environmental variance and thus reducing the estimate of heritability. Rapid and recent economic change in developing countries also creates different growth environments for children and their parents, thereby decreasing parent–child correlations and thus reducing the estimate of total variation attributable to genes. More recent studies of adult height in rapidly developing countries report h^2 estimates more similar to those that have typically been reported for the adult height of individuals in developed countries. For example, Li et al. [23] reported an estimated h^2 of 0.647 (±0.122) in a Chinese population, and Arya et al. [13] reported an h^2 of 0.72 in an Indian population. Whether lower heritability estimates are due to a larger absolute contribution of environmental variance, increasing the total variability and therefore lowering the h^2 estimate, or to a lower absolute contribution of genetic variance (in the case of an equal total variation) is not easily verifiable, since most studies do not report the variance components, but only the relative contribution of genetic factors (h^2).

Weight. In general, genetic determination of normal body weight is somewhat lower than that of height. This is probably because body weight is more susceptible to environmental influences such as dietary intake and energy expenditure. Estimates of the h^2 of body weight vary considerably from 0.20 to about 0.90. As for height, generally lower h^2 estimates for body weight in developing countries are probably due primarily to nutritional and disease stress and to differences in environmental conditions for children and their parents.

Patterns of growth. Analysis of genetic influences on patterns of growth requires longitudinal familial data so that individual growth curves can be fitted and parameters quantifying the timing and magnitude of growth over time derived. Such data are much rarer than cross-sectional data from unrelated children, but the data that do exist reveal many interesting findings. For example, differences between ethnic groups in patterns of growth have been found, indicating a role for genetic factors (and/or different environmental factors). Murata and Hibi [24] found an earlier occurrence of the adolescent growth spurt in Asian populations than in Caucasian populations. In meta-analyses of European and US growth studies, Hermanussen et al. [25], however, found that there were no major differences in growth patterns between populations in these countries and that the patterns have remained fairly similar during the twentieth century, despite a striking increase in mean body height, suggesting that individuals from populations with similar ethnic backgrounds are generally similar in their patterns of growth.

Lower sibling and twin correlations during puberty (for girls around 11 to 13 years and for boys around 13

to 15 years of age) are more marked for DZ twins than for MZ twins, suggesting that not only height but also the timing of the height spurt is under genetic control [26, 27]. Differences in the individual timing of the growth spurt make twins less similar when studied cross-sectionally at the same chronological ages during puberty.

Other studies have specifically estimated the h^2 of different growth curve parameters (see **table 2**). For example, in the Leuven Longitudinal Twin Study, the Preece–Baines model I was applied to semiannual measurements of stature in 99 pairs of twins aged 10 to 18 years [28]. Parameters of the pubertal growth spurt were derived, including the timing (i.e., age), the magnitude (i.e., velocity), and stature at the prepubertal nadir of growth (i.e., just prior to takeoff of the adolescent growth spurt) and at the peak of pubertal growth. In addition, the percentage of adult stature at both the prepubertal nadir of growth and the peak of pubertal growth and the predicted adult stature were derived. A model including additive genetic and specific environmental variance components in males and females best explained most of the variation observed in the growth parameters. For the timing and velocity of the adolescent growth spurt, no sex heterogeneity was observed, and the genetic (0.89 to 0.93) and specific environmental (0.07 to 0.11) contributions were equal in both sexes. The small sex differences could be explained by overall differences in total variance between the velocities of growth at takeoff and at PHV. Parameters pertaining to percentage of adult stature, or distance to grow from prepubertal takeoff or age at PHV, however, indicated a different set of genes to be active in males and females. Beunen et al. [28] reported a significant component of nongenetic familial resemblance (0.39 to 0.56) for the attained height at PHV and for predicted adult height. In the Fels Longitudinal Study, nontransmissible sibling resemblance was also present for age at PHV in boys and for stature at takeoff of pubertal growth and at PHV in girls [29].

Heritability estimates for the timing of the adolescent growth spurt in Swedish and Polish twins range from 0.49 to 0.76, except for PHV in Swedish girls [26, 30]. Within-pair growth curve similarities were higher in Swedish MZ twins ($r = 0.85$) than in DZ twins ($r = 0.54$) [31]. Intrapair correlations from MZ and DZ twins in the Wraclow twins and Swedish twins indicate lower heritability estimates for the timing of peak growth in weight and peak weight velocity, although substantial genetic factors seem to remain important [26, 32].

Towne et al. [33] used a simple three-parameter function to fit individual growth curves to serial recumbent length data from 569 infants aged from birth to 2 years from nuclear and extended families in the Fels Longitudinal Study. Substantial h^2 estimates of 0.83 for recumbent length at birth, 0.67 for rate of increase in length, and 0.78 for a parameter describing the curvilinear shape of growth in recumbent length from birth to 2 years were found. In addition, genotype-by-sex (GxS) interaction was indicated for the latter two growth parameters, suggesting that the genes influencing rate of growth and intrinsic rate of change in growth are influenced by the sex of the individual. In a subsequent multivariate quantitative genetic analysis of the pubertal growth spurt, Towne et al. [34] used a triple logistic model to fit individual growth curves to serial stature data from 471 Fels Longitudinal Study participants aged 2 to 22 years. Highly significant h^2 estimates were found for age at PHV (0.85), growth rate at PHV (0.61), and stature at age of PHV (0.96). Additive genetic correlations between these growth-spurt parameters were all lower than 1.0, indicating that the timing of the pubertal growth spurt, its magnitude, and attained stature at the time of PHV are controlled by genes with partial pleiotropic effects.

Longitudinal genetic model-fitting can also be applied to serial data to examine the nature of genetic and environmental effects over time during childhood growth (i.e., testing whether or not genetic or environmental influences remain constant or change during growth). For example, in a sample of Swedish female twins, Fischbein found significant and similar genetic contributions to weight over the delimited age range of 11.5 to 14 years of age and only modest evidence for age-specific genetic influences on growth [35]. Another approach is to fit latent growth-curve models to longitudinal data and estimate the genetic and environmental contributions to the underlying latent growth-curve parameters simultaneously with the time-point–specific variation not explained by the growth curve. Peeters et al. [36, 37]* tested several underlying growth curves to longitudinal data of height and body weight of the Leuven Longitudinal Twin Study. Variance and covariance between Preece-Baines I curve parameters (timing and tempo of growth) was best explained by shared additive genetic factors and unique environmental factors, with slightly higher values for height than for weight. Genetic variation in the underlying growth curve was the most important source of variance to explain the overall variation in longitudinal growth in height and weight (h^2 above 0.90 in height and above 0.80 in weight).

Unfortunately, longitudinal familial data from populations in developing countries are scarce. Sibling similarity in annual growth increments in a small sample of schoolchildren aged 6 to 13 years from a subsistence agricultural community in Oaxaca, Mexico, were negative or close to zero (44 brother–brother, 44 sister–sister, and 110 unlike–sex pairs). The lack of similarity between siblings in annual growth increments perhaps

* Peeters M, Thomis M, Maes HH, Loos R, Claessens AL, Vlietinck R, Beunen G. Structured latent growth curves applied to adolescent growth in stature and body weight. 2006. Behavior Genetics (unpublished).

TABLE 2. Heritabilities and family correlations of growth-curve parameters (timing and tempo) for height

	Stockholm [26][a]		Wroclaw [30][a]		Fels [29])[b]		Fels [34][a]		Leuven [28][a]	
Parameter	Boys	Girls	Boys	Girls	Parent–child	Sibs	Boys	Girls	Boys	Girls
Age at take-off			0.49		0.17	0.32			0.93	0.93
Velocity at take-off			0.61		0.26	0.35			0.90	0.90
Height at take-off			0.73						0.96	0.92
Age at PHV	0.74	0.64	0.74		0.22	0.35	0.85	0.85	0.92	0.92
PHV	0.56	0.00	0.76		n.s.	0.32	0.61	0.61	0.89	0.89
Height at PHV							0.96	0.96	0.39	0.94

PHV, peak height velocity; n.s., not stated
a. Heritability estimates.
b. Family correlations.

reflects age-specific variation in a genotype–environment interaction with chronic undernutrition, the important mediating environmental variable [38]. Many more such studies are needed, however, before any definitive statements can be made regarding cross-population differences in genetic influences on patterns of growth.

Skeletal maturation

Koniarek [39] determined skeletal age and skeletal maturation scores using the Tanner-Whitehouse (TW2) method in 55 pairs of male MZ and 55 pairs of male DZ twins, and 47 pairs of female MZ and 43 pairs of DZ twins, followed longitudinally in the Wroclaw twin study. The mean intrapair difference in skeletal maturity score was considerably lower for MZ twins of both sexes than for DZ twins across the age range of 7 to 18 years. Similarly, the standard deviation of the intrapair difference was also much smaller for MZ twins than for DZ twins. The highest intrapair differences were found in the period of 12 to 14 years in males and 9 to 13 years in females. The h^2 of skeletal age at the onset of menses in these female twins was estimated at 85% [40]. In a multivariate longitudinal analysis of Wroclaw Twin Study data, evidence was found for highly integrated genetic processes underlying both growth in stature and skeletal maturation, which have a different timing in boys and girls, relating to different effects of estrogen and testosterone/androgens in the stimulation of linear growth and acceleration or deceleration of bone maturation [41].

These findings are in concordance with those obtained in analyses of skeletal ages at the chronological ages of 3, 6, 9, 12, and 15 years, as determined by the Fels method [7], from 742 subjects, in 124 to 172 families in the Fels Longitudinal Study [42]. All available patterns of familial resemblance across relatives of varying degrees of relationships are taken into account to estimate the heritability of skeletal age at these five ages from late infancy to postpuberty. The h^2 of skeletal age was highest at 3 and 6 years of age (h^2 = ~1.00 and 0.97, respectively) and decreased to 0.48 by 15 years. The genetic correlations between skeletal ages closely positioned in time (e.g., at 3-year intervals) were high (> 0.84) and decreased with increasing time intervals (e.g., at 6-year intervals they ranged from 0.56 to 0.73, and at 9-year intervals they ranged from 0.30 to 0.37) and were lowest at the 12-year interval (0.16). These results indicate that although there are some genetic influences on skeletal maturation that appear to act throughout childhood, there are also some genetic influences that are more time-delimited in their action. Skeletal maturation is a complex process, and different biological phenomena occur at different stages of development. Early childhood is predominantly a time of bone ossification; in mid-childhood there are changes in bone shape and joint formation; and in puberty epiphyseal fusion begins to take place. Structural genes might be responsible for the appearance of bones and subsequent changes in their shape, whereas regulatory genes (e.g., hormones) may influence more the tempo of skeletal maturation.

At 10 and 13 years, the Tanner–Whitehouse II-20 scores in twins from the Leuven Longitudinal Twin Study (20 to 25 pairs of twins in five twin-by-sex zygosity groups) were highly determined by genetic factors. At 10 years, $h^2 = 0.92$ and $e^2 = 0.08$; at 13 years, $h^2 = 0.88$ and $e^2 = 0.12$. Sex differences in the variability of Tanner–Whitehouse II-20 scores at the chronological age of 13 years could be attributed to a general scalar effect [43].

In a recent preliminary analysis of skeletal maturation in Nepali children participating in the Jiri Growth Study, the h^2 of tempo of skeletal maturation from middle to late childhood (chronological ages 6 to 18 years) was high at 0.92 [44].

Sexual maturation

Age at menarche. Menarche is the hallmark maturational event of female puberty, and many studies have shown the timing of the onset of the menses to be significantly influenced by genetic factors (see Towne et al. [45] for a recent review). For example, MZ female twin correlations for age at menarche are high (0.65 to 0.90)

[40, 46], with DZ twin correlations ranging from 0.16 to 0.60. Sibling correlations are generally around 0.40 [47], and mother–daughter correlations generally range from 0.24 to 0.39 [47–51]. The mean difference in age at menarche is smallest for MZ twins (0.4 years), with increasing mean intrapair differences for DZ twins (0.8 years), siblings (1 year), and mother–daughter pairs (1.1 year). Some studies report high heritability estimates for this sexual maturation milestone ($h^2 = 0.95$), with evidence for dominance effects ($d^2 = 0.54$, $a^2 = 0.17$) [46], or shared genetic effects with skeletal maturity (h^2 unique to age at menarche = 0.44, h^2 shared with skeletal maturity = 0.53) [40]. More recent reports using prospective age at menarche data and larger kinship data report h^2 estimates around 0.50 [45].

Genital development and secondary sexual characteristics. Data from the Wroclaw Longitudinal Twin Study show a higher concordance between MZ males for genital developmental stages (G2 to G5) and pubic hair development stages (PH2 to PH5) than between DZ males [52]; similarly, MZ females are more concordant than DZ females for stages of breast and pubic hair development [53].

Linkage and association studies

To date, relatively few modern whole-genome scans searching for QTL influencing measures of normal growth and development have been conducted. Such linkage studies pose logistic and financial challenges, since familial data must be collected and extensive marker genotyping must be carried out. There are many more studies of associations between polymorphic markers in candidate genes and measures of growth and development. This is because data from related individuals are not necessarily needed, and only one or a few markers need to be genotyped. Unfortunately, the results of association studies are very often equivocal and contradictory. There are many possible reasons for this, the primary ones being small sample sizes, unaccounted-for population stratification, and the unpredictable nature of disequilibrium between the genotyped marker and the functional polymorphism influencing the trait of interest [54]. In the future, whole-genome association mapping studies using tens of thousands of genetic markers may be used to discover functional polymorphisms related to complex traits, including measures of human growth and development.

Linkage results

Stature. Most whole-genome linkage scans on growth and development measures have been done on adult stature. This is primarily because adult stature is often self-reported in large, questionnaire-based, genetically oriented population studies (e.g., families, twin registries), and because stature was one of the earliest phenotypes for which a large genetic component was observed (from twin and family studies). There are few linkage studies of height or weight phenotypes in the preadolescent and adolescent period, because there exist few studies of genetically related individuals in which sufficient data have been collected from them during their childhood.

A recent study reporting identification of QTLs for height in a Dutch sibling-pairs study also reviews linkage results for adult stature in other populations [55]. Adult stature, with a Gaussian distribution, is a multifactorial trait, and genetic influence on it is probably due to the cumulative effects of allelic variants at several loci. Significant linkage LOD scores (i.e., LOD scores > 3.0) and suggestive LOD scores (i.e., minimum LOD scores > 1.0) are spread at 92 locations over 21 chromosomes. Regions on chromosomes 3, 6, (harboring estrogen receptor alpha), 7, 12, 13, and 14 show the highest linkage peaks (LOD scores > 3.0). In a small sample of 79 sibling pairs, significant linkage ($p = .004$) was found between a dinucleotide marker in the dopamine 2 receptor gene and height in 79 sibling pairs aged 7 to 18 years [56].

Hirschhorn et al. [57] explored the differences in linkage results (especially lack of replication) from four different populations (Botnia, Finland; Finland; Southern Sweden; and Sagueney-Lac-St. Jean, Quebec). From simulation studies, they concluded that these differences could be due to sampling variation only, but that population-specific differences in the occurrence of rare or common alleles or differences in linkage disequilibrium patterns could also contribute to the different results. Population-specific interactions with other genes or environmental factors need to be taken into account when different linkage peaks are reported in different populations.

Göring et al. [58] conducted whole-genome scans using data from European-American, Mexican-American, European, and Nepali populations and found suggestive evidence of linkage of adult height to several genomic regions. The linkage signals were not, however, necessarily consistent across the study populations, suggesting that the importance of individual candidate loci may vary in different populations.

The major fraction of variation observed in the stature of healthy individuals is suggested as being determined by autosomal genes, and the contribution of sex-linked genes, if any, is small [59]. Sex chromosomes do, however, have a fundamental role in setting the stage for growth [60], as can be observed in patients with sex chromosome aberrations. Growth retardation is observed in Turner's syndrome (X-linked), whereas for GCY (growth control gene) on Yq, a 9-cm increase in adult height is found independently of gonadal sex steroids [61].

Weight and body-mass index. Because of the relevance of adiposity to type 2 diabetes, cardiovascular

disease, and other disorders related to overweight and obesity, many linkage analyses of overweight, obesity, and other measures of adiposity have been conducted in recent years (see Perusse et al. [62] for a recent review). Virtually all of these studies, however, have been conducted on adults or have included children with adults in the analyses. Moreover, the focus is often more on phenotypes at the higher end of the distribution of overweight and obesity and less on the full range of normal variability in body weight. This situation is likely to change in coming years, however, as specific childhood precursors of adult disease risks come under closer scrutiny. A significant part of this endeavor will include searching for genes influencing childhood normal weight and overweight and other adiposity traits.

For example, Arya et al. [63] recently conducted a whole-genome scan for QTL influencing birthweight in Mexican-Americans and non-Hispanic Whites. In a sample of Mexican-Americans from San Antonio, Texas, they found significant evidence of linkage of birthweight to a QTL on chromosome 6q (LOD = 3.72). In a sample of non-Hispanic whites participating in the Fels Longitudinal Study, strongly suggestive evidence of linkage of birthweight to a QTL at the same location on chromosome 6q also was found (LOD = 2.84). This study not only provides replication of a significant linkage finding for a QTL on chromosome 6q for birthweight, but also provides evidence that this chromosome 6q QTL influences the intrauterine growth of children from two different ethnic backgrounds. In 782 randomly ascertained white siblings (521 full-sib pairs and 39 half-sib pairs) in the Bogalusa Heart Study, linkage results from a genome scan on longitudinal analysis of body-mass index (BMI) changes from childhood to young adulthood (area-under-the-curve analysis) indicated suggestive signals for linkage with regions on chromosomes 1, 5, 7, 12, 13, and 18 [64]. Several of these regions have been found to be significantly linked to body weight or obesity measures in studies of adults.

Skeletal maturation. In the Fels Longitudinal Study, skeletal ages were determined for 1,069 children, aged 1 to 17 years, from 220 families (9,865 total assessments of skeletal age) by the use of the Fels method [7]. An initial set of 478 subjects was genotyped for approximately 400 autosomal markers. Variance components-based linkage analysis (SOLAR) [65] found consistent evidence of a QTL on chromosome 8q that influences skeletal age at chronological ages 2 to 10 years (LOD scores ranged from 0.96 to 3.16 at whole-year intervals over this age range). Suggestive evidence of linkage of skeletal age at various chronological ages to markers on other chromosomes also was found. This unique linkage study of longitudinal skeletal age data found evidence of specific QTL containing as yet unidentified genes that influence the tempo of normal skeletal maturation during different stages of childhood development.

In recent analysis of data collected from Nepali children participating in the Jiri Growth Study, Towne et al. [44] recently found significant linkage (LOD = 3.32) of the tempo of skeletal maturation during middle and late childhood (ages 6 to 18 years) to genetic markers on chromosome 3p. Interestingly, in a recent bivariate linkage analysis of second metacarpal cortical bone thickness and skeletal age in a set of 600 10-year-olds from the Fels Longitudinal Study, Duren et al. [66] found significant joint linkage of both traits to the same location on chromosome 3p. Together, these two studies provide some confirmatory evidence of a gene or genes on chromosome 3p that influence skeletal maturation and bone growth during middle childhood in children from two very different populations.

Results of association studies

Reports on associations of allelic variants with growth in height, weight, and maturity characteristics are mainly focused on variations in genes coding for growth hormones, growth hormone receptors, and bone metabolism–related genes (**table 3**). There is only limited information on associations of gene polymorphisms with indicators of the timing and tempo of growth or other maturity characteristics. Furthermore, as often observed with association studies, there is a lack of consistency in the results, with only limited replication of findings. For the purposes of this chapter, a literature search using MEDLINE was performed combining the terms "height, weight, maturity, menarche, sexual maturation" with the terms "association, polymorphism" and "childhood, adolescent, growth."

Height and growth curve characteristics. In 183 male and 131 female participants in the Fels Longitudinal Study, pubertal growth parameters were estimated by the triple logistic method as implemented in the AUXAL program [5]. In boys, ESRa *Pvu*II and *Xba*I polymorphisms were significantly associated with the timing of the prepubertal nadir of growth, height at the prepubertal nadir, and height at the age of PHV. For the *Pvu*II polymorphism, boys homozygous for the rarer allele (pp) were 0.6 years younger ($p = .0035$) and 5.7 cm shorter ($p = .0026$) at the prepubertal nadir than heterozygous boys (Pp) or boys homozygous for the more common allele (PP). They were also 0.6 years younger ($p = .0088$) and 5.4 cm shorter ($p = .0023$) at the peak of the pubertal growth spurt. As adults, pp genotype males were 4.8 cm shorter ($p = .01$) than those with the PP genotype. In girls these trends were generally evident but were not as statistically significant. For example, height at PHV in girls with the pp genotype differed by 2.0 cm from that in girls with the PP genotype ($p = .047$), and the mean adult height of

TABLE 3. Summary of association studies related to growth in height, weight, adiposity, age at menarche, and other maturity-related characteristics[a]

Gene	Polymorphism	Gene function	Phenotype	Effect	Study sample	Reference
ACP1 (acid phosphatase-1 gene)	ACP1 A and BA	Modulation of flavo-enzyme activity and energy metabolism	BMI	Association with BMI	Italian obese children	Lucarini et al. [73]
ADRB3 (beta-3-adrenergic receptor gene)	Trp64Arg	Regulation of lipolysis and thermogenesis in adipose tissue	BMI (age, sex adjusted)	Arg/Arg and Trp/Arg significant Higher BMI than Trp/Trp (19.4 ± 3.6 vs 18.9 ± 3.2, $p = .02$)	291 Japanese boys 262 Japanese girls (9–15 yr)	Endo et al. [74]
			Weight/BMI (age, sex adjusted	No association	311 Chinese (8-11 y)	Xinli et al. [75]
			Weight/BMI gain after dietary intervention in obese	Lower increases in weight and BMI in obese children without the mutation compared to control group	36 obese Chinese	Xinli et al. [75]
			Obesity/underweight	No significant allele frequency differences	296 obese/134 healthy underweight (Germany)	Tafel et al. [76]
ADRB1 (beta-1-adrenergic receptor gene)	Gly49 Ser Arg389Gly	Regulation of lipolysis and thermogenesis in adipose tissue	Obesity/underweight	No significant allele frequency differences	296 obese/134 healthy underweight (Germany)	Tafel et al. [76]
ADRB2 (beta-2-adrenergic receptor gene)	Arg16Gly Gln27Glu	Regulation of lipolysis and thermogenesis in adipose tissue	Obesity/underweight	No significant allele frequency differences	296 obese/134 healthy underweight (Germany)	Tafel et al. [76]
AR (androgen receptor)	Exon 1 GGC repeat (16 repeats/ other alleles)	Binds to testosterone, increased transcription of androgen-sensitive genes	Age at menarche	16/16 GGC repeats genotype significant earlier age at menarche (12.07 ± 1.5 yr) compared to other genotypes (12.68 ± 1.64 yr)	164 non-Hispanic white	Comings et al. [92]
COMT (cytochol-*O*-methyltransferase)	*Hsp*II *92* (H/L)	Inactivation of the reactive metabolites in estrogen pathway	Age at menarche	No association	Japan ($n = 317$)	Gorai et al. [85]
CYP17 (cytochrome P450 gene family member)	*Msp*AI (A1 low activity/A2 high activity)	Estrogen (E2) biosynthesis	Age at menarche	No association	583, controlled for ethnic group and year of birth	Lai et al. [87]

Gene	Polymorphism	Gene function	Phenotype	Effect	Study sample	Reference
				A2/A2, and A1/A2 (13.6 ± 1.2 yr) have earlier age at menarche (A1/A1, 14.1 ± 1.3 yr)	Japan (n = 317)	Gorai et al. [85]
				A1/A1 later menarche	Women without breast cancer	Feigelson et al. [86]
			Onset of breast development (T2B)	No association	137 (9.5 yr), controlled for ethnic group and somatic differences	Kadlubar et al. [94]
CYP1A1 (cytochrome P450 gene family member)	*Msp*I (wt/vt)	Estrogen hydroxylation	Age at menarche	No association	Japan (n = 317)	Gorai et al. [85]
CYP1A2 (cytochrome P450 gene family member)	*Eco*57I **1F variant*	Estrogen hydroxylation (catalyzing 2- and 4-hydroxylation of estradiol and estrone)	Age at menarche	No association	583, controlled for ethnic group and year of birth	Lai et al. [87]
			Onset of breast development (T2B)	No association	137 (9.5 yr), controlled for ethnic group and somatic differences	Kadlubar et al. [94]
CYP1B1 (cytochrome P450 gene family member)	**3 variant* *Eco*57I (Val/Lue)	Estrogen hydroxylation (catalyzing 4-hydroxyestradiol)	Age at menarche	No association	583, controlled for ethnic group and year of birth	Lai et al. [87]
			Onset of breast development (T2B)	No association	137 (9.5 yr), controlled for ethnic group and somatic differences	Kadlubar et al. [94]
CYP3A4 (cytochrome P450 gene family member)	*1B variant (W/V)	Deactivation of testosterone	Age at menarche	No association	583, controlled for ethnic group and year of birth	Lai et al. [87]
			Onset of breast development (T2B)	*1B: 3.2 adjusted OR to have T2B (or +) compared with T1B	137 (9.5 yr), controlled for ethnic group and somatic differences	Kadlubar et al. [94]
CYP3A5 (cytochrome P450 gene family member)	*1 variant	Confers CYP3A activity in liver and kidney	Onset of breast development (T2B)	No association	137 (9.5 yr), controlled for ethnic group and somatic differences	Kadlubar et al. [94]

continued

TABLE 3. Summary of association studies related to growth in height, weight, adiposity, age at menarche, and other maturity-related characteristics[a] *(continued)*

Gene	Polymorphism	Gene function	Phenotype	Effect	Study sample	Reference
DRD2 (dopamine D2 receptor gene)	Polymorphism in promoter region	Neurotransmitter receptor	Height	$p = .009$, paired t-test, in the sib pairs; $p = .006$, ANOVA, in the adults	79 sib pairs (7–18 yr) 135 adults (Japan)	Arinami et al. [56]
ERα (estrogen receptor alpha)	*Pvu*II (P/p) *Xba*I (X/x)	Binding to estrogen	Age at menarche	PPXX (homozygous haplotype) later menarche: 13.43 ± 1.18 vs 12.76 ± 1.25 yr compared with other haplotypes	Greece, rural ($n = 145$)	Stavrou et al. [83]
				No association	Netherlands ($n = 90$)	Boot et al. [84]
				No association	Japan ($n = 317$)	Gorai et al. [85]
			Age, height, and growth rate at prepubertal nadir (AUXAL)	No association with *Xba*I polymorphism In boys/girls (*Pvu*II pp vs. Pp or PP): pp: 0.6/trend yr younger at prepubertal nadir pp: 5.7/trend cm shorter at prepubertal nadir	183 male, 131 female (whites, Fels Study, Ohio, USA)	Parks et al. [67]
			Age, height and growth rate at PHV	pp: 0.6/trend yr younger at age of PHV pp: 5.4/2.0 cm shorter at age of PHV		Parks et al. [67]
			Adult height	Pp: 4.8/2.5 cm shorter at adult height		Parks et al. [67]
			Height at 7 yr	no association	56 males, 68 females, white (7 yr)	Tao et al. [69]
			Height at 16.7 yr	Trend 4 cm shorter at near-adult height	90 adolescent whites	Lorentzon et al. [68]

Gene	Polymorphism	Gene function	Phenotype	Effect	Study sample	Reference
GnRH1 (gonadotropin-releasing hormone)	Sequencing and haplotype study	Stimulates secretion of luteinizing hormone and follicle-stimulating hormone trough binding with GnRHR	Late age at menarche vs early age at menarche	No clear associations or TDT with haplotypes or htSNPs in haplotype block of GnRH1	81 late-onset puberty (non-medical) + parents (US) (TDT) 44 late-onset puberty + parents (UK) 506 females (menarche < 11yr vs. +15 yr)	Sedlmeyer et al. [93]
GnRHR (gonadotropin-releasing hormone receptor gene	5′ region 10-bp insertion at 1187 position (I/D)	Transduces signals from GNRH and modulates the synthesis and secretion of luteinizing hormone and follicle-stimulating hormone	Age at menarche	No association	196 Japanese (18–20 yr)	Nanao et al. [91]
	Sequencing and haplotype study		Late age at menarche vs early age at menarche	Homozygous hCV3145733 were 1.85 times more likely to have a late menarche Haplotype 4 in block 1 of GnRHR was associated with reduced risk of late menarche (OR, 0.52) No replication in the different datasets	81 late-onset puberty (non-medical) + parents (US) (TDT) 44 late-onset puberty + parents (UK) 506 females (menarche < 11yr vs +15 yr)	Sedlmeyer et al. [93]
GRL (glucocorticoid receptor)	GRL IVS2-*Bcl*I	Receptor for cortisol, corticosterone, exerts orexigenic and antithermogenic effects	Increase in adiposity (12-yr follow-up), sum of skinfolds	4.5/2.3 female carriers had more than 2× increase in adiposity compared with 4.5/4.5 and 2.3/2.3 genotype subgroups	n = 83 (14.3 yr at baseline)	Tremblay et al. [82]
HSD11B1 (11-beta-hydroxysteroid dehydrogenase type 1)	ins4436A	Glycoprotein enzyme, cortisone to cortisol conversion in visceral adipose tissue	BMI-SD (age, sex, race, height corrected) Waist circumference Waist-to-hip ratio	Greatest BMI-SD for ins4436A homozygotes. Homozygotes also had greater waist circumference, waist-to-hip ratio, and insulin resistance indices than heterozygote or wild-type children (all p<.05)	160 normal BMI (-2/+2 SD) US children 103 overweight BMI (>2 SD) US children	Gelernter-Yaniv et al. [77]

continued

TABLE 3. Summary of association studies related to growth in height, weight, adiposity, age at menarche, and other maturity-related characteristics[a] *(continued)*

Gene	Polymorphism	Gene function	Phenotype	Effect	Study sample	Reference
LEP (leptin)	D7S1875 (<280 bp vs ≥280 bp)	Suppression of neuropeptide Y induces onset of puberty through release of inhibition of pituitary-gonadotropin axis	Age at menarche	Association-crossover effect depending on maternal age at birth. ≥280 bp: age at menarche 0.75 yr later than <280 bp if maternal age ≥ 30 yr. 0.75 yr earlier when maternal age <30 yr	183 whites	Comings et al. [90]
LH (luteinizing hormone-beta gene)	Trp8Arg and Ile15Thr	Gonadotropic hormone from anterior pituitary with signaling action to ovarian and testes development	Testicular volume Height Growth rate	Homozygote/heterozygote mutants: smaller testicular volumes ($p < .03$), shorter ($p < .02$), slower growth rates ($p < .04$)	49 Finnish boys, 11.7 yr, followed for 3 yr every 3 mo	Raivio et al. [71]
MC4R (melanocortin-4-receptor)	V103I, I125L, Y35X, D37V, S30F, S127L, R165W, L211fsX216	Receptor in brain, food intake regulation	Obese/nonobese	Association and significant transmission disequilibrium test	808 extremely obese German children (13 yr), 502 with parents (TDT), 327 nonobese	Hinney et al. [79]
NOS3 (endothelial nitric oxide synthase)	Glu298Asp T-786C	Neuroendocrine function in reproduction and ovulation	Age at menarche	No association	87 whites, postmenopausal	Worda et al. [88]
	Tandem repeat in intron 4 (A/B)			No association	91 whites, postmenopausal	Worda et al. [88]
NPY (preproneuropeptide Y gene)	Leu7Pro	Neurotransmitter, favors energy storage	Birthweight	Pro/Leu + Pro/Pro boys 193 g higher birthweight than Leu/Leu homozygotes ($p = .03$), not in girls	688 7-yr-old children, Finland	Karvonen et al. [80]
SHP (small heterodimer partner gene)	R34G, R36C G171A, –195CTGAdel	Inhibits transcriptional activity of hepatocyte nuclear factor-4alpha	Birthweight (corrected for. gestational age)	–195CTGAdel, D/D and W/D lower birthweight than W/W	GOOS, ALSPAC (7 yr), Ely study	Hung et al. [81]

Gene	Polymorphism	Gene function	Phenotype	Effect	Study sample	Reference
			BMI	ALSPAC: G171A carriers (7 yr) have greater BMI (19.4 vs 16.1) and waist circumference than noncarriers		Hung et al. [81]
VDR (vitamin D receptor)	*Apa*I (A/a)	? function in reproductive organs, growth plate	Age at menarche	Earlier menarche in Aa (12.1 ± 1.0 yr) than in aa (12.5 yr)	Japan (n = 120, 18–19 yr)	Kitagawa et al. [89]
	*Taq*I (T/t)		Height, weight at 7 yr	No association in boys TT girls 4.1 cm taller than Tt or tt TT girls 3.9 kg heavier than Tt or tt	56 male, 68 female whites (7 yr)	Tao et al. [69]

wt, wild type; vt, variant type
a. Genes are listed in alphabetical order of their symbols

females with the pp genotype was 2.5 cm shorter than that of females with the PP genotype. There was no effect of the *Xba*I polymorphism in ESRa on puberty-related growth parameters in either sex [67]. Lorentzon et al. [68] also had observed a similar trend for adolescent Caucasian boys with the ESRa *Pvu*II pp genotype to be shorter than PP boys at 16.9 ± 0.3 years, and they observed no effects of the XbaI genotype. However, no effect of the ESRa *Pvu*II genotype on height or weight was found in 7-year-old Caucasians in a small study by Tao et al. [69].

Vitamin D receptor (VDR) polymorphisms have been studied extensively in relation to bone mineral content and density in relation to osteoporosis and bone accretion during growth. Some studies have related variation in the VDR gene to characteristics of preadolescent growth and found that associations with Bone mineral density (BMD)/bone mineral content (BMC) disappeared after correction for size differences. In the sample studied by Tao et al. [69], females homozygous for a *Taq*I polymorphism in the vitamin D receptor gene (TT) were 4.1 cm taller and 3.9 kg heavier than those with the Tt or tt genotype.

Possible interactions between estradiol receptor gene (ER*Pvu*II) and VDR (*Bsm*I, BB or bb) polymorphisms have been studied by Suarez et al. [70] in 161 healthy Caucasian full-term babies. There was a lack of association between ER polymorphisms and body weight in boys and girls, body length in girls, or body length in boys with a bb genotype. However, the ER polymorphism and body length were significantly associated in BB boys, with a smaller length at birth and at age 10 months (but not at 2 years of age) for those with a BBpp genotype.

An association between stature and a putative functional polymorphism in the promoter region of the dopamine 2 receptor gene (DRD2) was examined in the 79 sibling pairs aged 7 to 18 years and in 125 unrelated male Japanese adults. The association with stature (p = .009, paired t-test, in the sibling pairs; p = .006, ANOVA, in the adults) was suggestive of a role of the DRD2 promoter polymorphism in stature [56].

Raivio et al. [71] studied two point mutations in the ß-subunit of the luteinizing hormone-beta gene (LH-ß) in a group of 49 healthy boys followed from a mean age of 11.7 years at 3-month intervals. Thirty-six boys (74%) were homozygous for the wild-type LH-ß allele, 12 (24%) were heterozygous carriers of the variant allele, and 1 (2%) was homozygous for the variant allele. Boys with the variant allele were shorter (p < .02), had slower growth rates (p < .04), and had lower serum insulin-like growth factor I–binding protein-3 levels (p < .03) than boys homozygous for the wild-type LH-ß allele. In boys with delayed onset of puberty, the frequency of the variant LH-ß allele did not differ from that in the reference population, indicating that the variant allele is not associated with conditions due to

disturbed control of the reactivation of GnRH secretion. During the progression of puberty, the variant LH-ß allele may be less active in stimulating testicular growth than the wild-type LH-ß allele. Thus, the gene may affect tempo, contributing to the wide normal variation in pubertal progression in healthy boys. Similar findings were observed in US boys and girls in the Fels Longitudinal Study, but the association between the variant LH-ß allele and height during puberty was not as pronounced as in the sample of Finnish boys [72].

Weight. As mentioned earlier, identifying specific genes influencing weight, overweight or obesity, and other measures of adiposity is of relevance for understanding genetic influences on adiposity-related measures of diabetes and cardiovascular disease risks, both in childhood and during adulthood. Although greater attention should be paid to genetic variation in normal weight for developing an International Growth Standard for Preadolescent and Adolescent Children, several studies have examined associations between polymorphisms in candidate genes and measures of body mass and adiposity in overweight or obese children, and these are of interest given the worldwide increase in obesity in both children and adults. For example, in a sample of Italian children, Lucarini et al. [73] found an association between BMI and a marker in the acid phosphatase-1 (ACP1) gene on chromosome 2p. In a sample of Japanese children, Endo et al. [74] found an association between BMI and a polymorphism in the beta-3-adrenergic receptor (ADRB3) gene on chromosome 8p. In a Chinese sample, Xinli et al. [75] found an association between body weight in obese children and a polymorphism in the ADRB3 gene. In a sample of German children, however, no associations between obesity and polymorphisms in the ADBR1, ADBR2, or ADBR3 genes were found [76]. In a sample of US children, Gelernter-Yaniv et al. [77] found an association between a polymorphism in the 11-beta-hydroxysteroid dehydrogenase type 1 (HSD11B1) gene on chromosome 1q and BMI, waist circumference, and waist–hip ratio. In a German study, Roth et al. [78] reported transmission disequilibrium and sequence variants of the leptin receptor (LEPR) gene on chromosome 1p in a sample of extremely obese children. Hinney et al. [79] found evidence of association between polymorphisms in the melanocortin-4-receptor (MC4R) gene on chromosome 18q and extreme obesity in a sample of German children. In a Finnish study, Karvonen et al. [80] found an association between a polymorphism in the preproneuropeptide Y (NPY) gene on chromosome 7p and birthweight. Hung et al. [81] found evidence of associations between polymorphisms in the small heterodimer partner (SHP) gene on chromosome 1p and birthweight and adiposity during childhood. Tremblay et al. [82] found an association between a polymorphism in the glucocorticoid receptor (GRL) gene on chromosome 5p and increase in adiposity in girls from adolescence to young adulthood.

Age at menarche. Only a few polymorphisms have been associated with age at menarche, and there is a lack of replication in different samples or populations. Because early age at menarche is a risk factor for the development of breast cancer, many genes involved in breast cancer have been studied, and some are related to age at menarche (however, these breast cancer gene polymorphisms have not been included systematically in this review).

In addition to pubertal growth, as discussed earlier, associations between the *Pvu*II (P/p) and *Xba*I (X/x) polymorphisms in the ESRa gene and age at menarche have also been studied. In a closed rural community in northwestern Greece, the *Pvu*II and *Xba*I haplotype was significantly related to a later age at menarche (homozygous PPXX versus all other haplotypes: 13.43 ± 1.18 years versus 12.76 ± 1.25 years) [83]. However, in a study of 90 girls (mean age, 15.6 years), the *Pvu*II and *Xba*I haplotype was not related to age of menarche [84]. Furthermore, no association was found in 317 Japanese women [85]. These authors, however, report a significant association of age at menarche with the *Msp*AI polymorphism in the CYP17 gene, one of the estrogen-metabolizing genes (estrogen biosynthesis, P450c17a). Women with higher CYP17 activity (A2/A2 and A1/A2 genotypes) tended to have an earlier age at menarche (13.6 ± 1.2 years) than women with lower CYP17 activity (A1/A1 genotype; 14.1 ± 1.3 years) [85]. Similar findings for a later age at menarche for A1/A1 genotypes were found in women without breast cancer [86]. Polymorphisms in the hydroxylation (cytochrome P4501A; CYP1A1) gene and the gene regulating inactivation of the reactive metabolites (catechol-*O*-methyltransferase; COMT) were not associated with age at menarche in these Japanese women [85]. In a mixed ethnic sample of 583 women, however, age at menarche was not significantly related to the CYP17 polymorphism nor to polymorphic variants of the CYP3A4, CYP1B1, and CYP1A2 genes [87]. NOS3 gene variants (Glu298Asp, T-786C, repeat in intron 4) were studied in 87 to 91 Austrian women, although no significant association with age at menarche could be found [88]. The vitamin D receptor gene VDR (*Apa*I polymorphism) was significantly related to an earlier age at menarche (12.1 versus 12.5 years) in 120 Japanese 18- to 19-year-old females [89].

Recently, more evidence has been found for the role of the leptin gene in the regulation of onset of puberty. Leptin gene variation (LEP D7S1875, < 280 bp versus ≥ 280 bp) has been shown to have opposite effects on age at menarche, depending on maternal age at the time of delivery of the girls under study [90]. In mothers giving birth to their daughters before the age of 30 years, the longer allele of the LEP D7S1875 polymorphism was associated with an earlier age at menarche in the daughters in comparison with heterozygotes

and homozygotes of the short alleles. In contrast, for maternal ages ≥ 30 years, daughters with the ≥ 280 bp/≥ 280 bp genotype had a later age at menarche. A 10-bp insertion in the 5' end of the gonadotropin-releasing hormone receptor (GnRHR) gene was not associated with age at menarche in 196 Japanese girls aged 18 to 20 years [91]. Within the androgen receptor gene, a GGC repeat polymorphism (16 repeats versus polymorphisms with other repeat numbers) has been associated with an earlier age at menarche for 16 GGC repeat carriers as compared with genotypes with other repeat numbers [92].

A recent report by Sedlmeyer et al. [93] searched for sequence variation in both GnRHR and GnRH1 in males and females with late maturity onset (children with late maturity and their parents (UK and US sample) and in an association sample of 506 females from the US Multiethnic Cohort Study of Diet and Cancer. Haplotype block analyses were performed, as well as htSNPs analysis within the detected haplotype blocks. The authors concluded that the studied haplotypes and htSNPs within GnRHR and GnRH1 seemed to have only limited influence on the timing of puberty.

Breast development. The onset of breast development (≥ T2B stage versus T1B) could be predicted from CYP3A4*B1 variant genotypes in 9.5-year-old girls [94]; the B1 variant is associated with the onset of T2B. The onset of breast development was not associated with gene variants in other estrogen-metabolizing genes (CYP17, CYP1A1, CYP1A2, CYP1B1, and CYP3A5) [94].

Testicular volume. In their study of association between a variant of the LH-ß allele and pubertal growth of Finnish boys discussed earlier, Raivio et al. [71] also found that boys with the variant LH-ß allele had smaller testicular volumes ($p < .03$) than boys homozygous for the wild type LH-ß allele.

As the above discussion has pointed out, association studies have tested the significance of gene sequence variants in several genes in relationship to height, weight, and other maturity-related phenotypes. However, studies on children and adolescents in developing countries are lacking, and there is limited replication of findings. The degree of replication might even be underestimated because of positive publication bias; negative association results (i.e., showing lack of association) might not be published as often as positive association findings. Therefore, we do not have sufficient information to state that there are large population differences in the effects of specific gene variants, and whether this might be due to differences in the amount of the gene effect, differences in population-specific allele frequencies, or the contribution of different genes.

Conclusions

In the wake of the profound advances made in molecular and statistical genetic techniques and methods over the last two decades, advances that continue to be made at a rapid pace, the genetic architecture of complex processes such as those that constitute growth and development will become better elucidated in the coming years. Much of this better understanding of genetic influences on growth and maturation will be motivated by increasingly recognized associations between patterns of growth and the development of disease risks during adulthood. As alluded to in the forgoing sections of this chapter, there are two particular areas of research on the nature of genetic influences on growth and maturation that pertain directly to the broader issue of the establishment of an International Growth Standard for Preadolescent and Adolescent Children. These are population differences in major gene effects on growth and maturation, and the effect of gene–environment interactions on growth and maturation.

Population differences in major gene effects on growth

As yet there are too few cross-population genetic epidemiologic studies of growth and maturation to arrive at any definitive conclusions on this matter. Nonetheless, populations differ in gene frequencies, and it is reasonable to hypothesize that there are differences across populations in the frequencies of specific alleles that influence growth and development. This has implications for the relative importance of such specific genetic influences on growth and development in different populations. Allelic variation in a particular growth-related gene may explain a significant part of the observed variation in a measure of growth in one population but not necessarily in another. This does not mean that the fundamental genetic architecture of growth and development differs across populations, but that the relative importance of specific genes and their allelic variants might well differ across populations. Furthermore, the issue of nonreplication should be judged carefully. As mentioned earlier, in a series of simulations, Hirschhorn et al. [57] found that lack of replication in linkage studies across populations (for a true QTL explaining 20% of the trait variance) could be due simply to sampling variation.

Effects of gene–gene and gene–environment interactions

As stated earlier, gene–environment interaction effects are not easily studied, and the nature of gene–environment interaction effects across populations are complicated and difficult to assess. At present, there is very little knowledge about the effects of gene–environment interactions on height, weight, and maturity-related characteristics in children and adolescents. Large population-based studies in different areas of the world will

be needed to study the interaction of an individual's genotype with environmental factors, which can be beneficial (e.g., nutritional balance, physically active lifestyle) or harmful (e.g., nutritional imbalance, disease and illnesses, war, sedentary lifestyle) to the growing child. Because environmental factors, as well as the frequencies of pertinent alleles, might differ greatly between populations, the impact of gene–environment interaction on the observed variation in the growth and maturation of children across populations can be hypothesized to be significant. Knowledge of gene variants found to be protective against harmful environmental factors can be highly informative for the detection of children at risk for their individual response to those environmental factors (e.g., at risk for the development of obesity).

Implications for an international growth standard for preadolescent and adolescent children

The preceding discussion makes it clear that there is much that is as yet unknown regarding the nature of cross-population differences in genetic influences on growth and maturation. This gap in knowledge necessitates that caution be taken in the consideration of an International Growth Standard for Preadolescent and Adolescent Children, but it does not necessarily preclude the establishment of such a reference. Nonetheless, to reiterate two related points:

» Genetics plays a leading role in explaining individual variation within populations in growth and development traits, but there might be population differences in the genetic regulation of growth and development that have yet to be fully examined.
» It can reasonably be hypothesized that although the fundamental genetic underpinnings of the growth and development of children worldwide may be essentially the same, there may be differences between populations in the nature of gene-by-environment interactions.

As with any population-specific growth reference, the development of a single international growth standard for preadolescent and adolescent children will need to reflect both average growth and normal variation in growth. This should not pose an insurmountable problem, since most of the observed differences in complex quantitative traits such as measures of growth and development are probably contained within any sufficiently large group of subjects from within a population [95, 96]. Thus, the challenges are to determine what data should contribute to an International Growth Standard for Preadolescent and Adolescent Children, determine the form that this standard should take, and finally to determine whether the form that an international growth standard for preadolescent and adolescent children ultimately takes will prove useful in day-to-day application in different parts of the world.

As environmental sources of variation that negatively impact growth and maturation (e.g., disease and malnutrition) are ameliorated in developing nations, it will become somewhat easier to evaluate the nature of population variation in genetic influences on growth and maturation. It is reasonable to hypothesize that population differences in measures of growth and maturation may become smaller in the coming decades, but at the same time it is also reasonable to hypothesize that a genetic basis to population differences in measures of growth and development will remain, and that these will become primary sources of interpopulation differences in measures of growth and maturation. Such population differences will need to be accommodated in an International Growth Standard for Preadolescent and Adolescent Children.

One option for an International Growth Standard for Preadolescent and Adolescent Children, then, would be to create "median" values in growth for children and adolescents from many populations worldwide who are growing up in more or less environmentally ideal, or at least suitable, situations (e.g., balanced diet, regular exercise, peaceful home and social life, lack of major illnesses, lack of pollution, etc.). A list of known growth disorders with a genetic etiology also would need to be used as exclusion criteria. Overall variation in stature and other measures from such data would be almost entirely a reflection of population-specific genetic variation, population-specific environmental variation, and population-specific gene–environment interactions, which individually or collectively may or may not be large contributors to mean interpopulation differences in measures of growth.

In the application of an International Growth Standard for Preadolescent and Adolescent Children, there may need to be a paradigm shift among researchers and clinicians from thinking of specific height-for-age values in terms of percentiles to thinking in terms of a percentile ranking that consists of a range of potential values. Of course, if the range of values at any particular age becomes too large, then the utility of such a growth reference is compromised.

In conclusion, this review has summarized the current general understanding of genetic influences on within- and between-population variation in growth and development. As was discussed, almost all measures of growth and development have substantial and significant genetic underpinnings. But although major progress has been made in both molecular genetic and statistical genetic techniques over the last quarter-century, many questions remain unanswered and need further study. There is a need for more genetic epidemiologic studies of the growth and development of children from non-Western populations in order to better understand population differences in the genetic regulation of processes of growth and maturation, as well as to study specific gene-environment interactions and their effects on growth and development. A

variety of research designs can be used to contribute to this knowledge. An international growth standard for preadolescent and adolescent children will need to be particularly mindful of potentially significant differences between populations in the genetic regulation of growth and development that may exist even after detrimental environmental factors have been ameliorated.

References

1. Ulijaszek S. Environmental influences on preadolescent and adolescent growth in weight and height. Food Nutr Bull 2006; 27(suppl):S279–94.
2. Beunen G, Rogol AD, Malina RM. Indicators of biological maturation and secular changes in biological maturation. Food Nutr Bull 2006; 27(suppl):S244-56.
3. Preece MA, Baines MJ. A new family of mathematical models describing the human growth curve. Ann Hum Biol 1978;5:1–24.
4. Bock RD, Thissen DM. Statistical problems of fitting individual growth curves. In: Johnston FE, Roche AF, Susanne C, eds. Human physical growth and maturation: Methodologies and factors. New York: Plenum Press, 1980:265–90.
5. Bock RD, du Toit SHC, Thissen D. AUXAL: Auxological analysis of longitudinal measurements of human stature (version 3). Lincolnwood, Ill, USA: Scientific Software International, 2005.
6. Tanner JM, Whitehouse RH, Cameron N, Marshall WA, Healy MJR, Goldstein H. Assessment of skeletal maturity and prediction of adult height (TW2 method). London: Academic Press, 1975.
7. Roche AF, Chumlea WMC, Thissen D. Assessing the skeletal maturity of the hand-wrist: Fels Method. Springfield, Ill, USA: Charles C. Thomas, 1988.
8. Tanner JM. Growth at adolescence, 2nd ed. Oxford, UK: Blackwell Science, 1962.
9. Neale MC, Cardon LR. Methodology for genetic studies of twins and families. Dordrecht: Kluwer, 1992.
10. Lynch M, Walsh B. Genetics and analysis of quantitative traits. Sunderland, Mass, USA: Sinauer Associates, 1998.
11. Falconer DS, Mackay TFC. Introduction to quantitative genetics. Harlow, UK: Longman, 1996.
12. Silventoinen K. Determinants of variation in adult body height. J Biosoc Sci 2003;35:263–85.
13. Arya R, Duggirala R, Comuzzie AG, Puppala S, Modem S, Busi BR, Crawford MH. Heritability of anthropometric phenotypes in caste populations of Visakhapatnam, India. Hum Biol 2002;74:325–44.
14. Luke A, Guo X, Adeyemo AA, Wilks R, Forrester T, Lowe W Jr, Comuzzie AG, Martin LJ, Zhu X, Rotimi CN, Cooper RS. Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. Int J Obes Relat Metab Disord 2001;25:1034–41.
15. Towne B, Demerath EW, Czerwinski SA. The genetic epidemiology of growth and development. In: Cameron N, ed. Human growth and development. Amsterdam: Academic Press, 2002.
16. Williams-Blangero S, Towne B, Subedi J, Blangero J. Genetic studies of the Jirels: From population genetics to genetic epidemiology. In: Subedi J, ed. Halfway to the mountain: the Jirels of Eastern Nepal. Kathmandu: Tribhuvan University Press, 2002:170–87.
17. Sham PC. Statistics in human genetics. London: Arnold, 1998.
18. Beunen G, Thomis M. Gene Powered? Where to go from heritability (h^2) in muscle strength and power? Exerc Sport Sci Rev 2004;32:148–54.
19. Palmert MR, Hirschhorn JN. Genetic approaches to stature, pubertal timing, and other complex traits. Mol Genet Metab 2003;80:1–10.
20. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, De Lange M, Harris JR, Hjelmborg JV, Luciano M, Martin NG, Mortensen J, Nistico L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemsen G, Kaprio J. Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res 2003;6:399–408.
21. Maes HH. Univariate and multivariate genetic analysis of physical characteristics of twins and parents. Leuven, Belgium: Katholieke Universiteit Leuven, 1992. PhD Thesis.
22. Mueller WH. Transient environmental changes and age-limited genes as causes of variation in sib-sib and parent-offspring correlations. Ann Hum Biol 1978;5:395–8.
23. Li MX, Liu PY, Li YM, Qin YJ, Liu YZ, Deng HW. A major gene model of adult height is suggested in Chinese. J Hum Genet 2004;49:148–53.
24. Murata M, Hibi I. Nutrition and the secular trend of growth. Horm Res 1992;38(suppl 1):89–96.
25. Hermanussen M, Thiel C, von Buren E, Rol de Lama MA, Perez Romero A, Ariznaverreta Ruiz C, Burmeister J, Tresguerres JA. Micro and macro perspectives in auxology: findings and considerations upon the variability of short term and individual growth and the stability of population derived parameters. Ann Hum Biol 1998;25:359–85.
26. Fischbein S. Onset of puberty in MZ and DZ twins. Acta Genet Med Gemellol (Roma) 1977;26:151–8.
27. Sharma JC. The genetic contribution to pubertal growth and development studied by longitudinal growth data on twins. Ann Hum Biol 1983;10:163–71.
28. Beunen G, Thomis M, Maes HH, Loos R, Malina RM, Claessens AL, Vlietinck R. Genetic variance of adolescent growth in stature. Ann Hum Biol 2000;27:173–86.
29. Byard PJ, Guo S, Roche AF. Family resemblance for Preece-Baines growth curve parameters in the Fels Longitudinal Growth Study. Am J Hum Biol 1993;5:151–7.
30. Hauspie RC, Bergman P, Bielicki T, Susanne C. Genetic variance in the pattern of the growth curve for height: a longitudinal analysis of male twins. Ann Hum Biol 1994;21:347–62.
31. Fischbein S, Nordqvist T. Profile comparisons of physical growth for monozygotic and dizygotic twin pairs. Ann Hum Biol 1978;5:321–8.
32. Bergman P. The problem of genetic determination of growth at adolescence. Materiały i Prace Antropologic-

zne 1988;108:165–216.
33. Towne B, Guo S, Roche AF, Siervogel RM. Genetic analysis of patterns of growth in infant recumbent length. Hum Biol 1993;65:977–89.
34. Towne B, Parks JS, Guo S, Siervogel R. Quantitative genetic-analysis of associations between pubertal growth-pattern parameters (abstract). Am J Hum Genet 1995; 57:989.
35. Fischbein S, Molenaar PC, Boomsma DI. Simultaneous genetic analysis of longitudinal means and covarianc structure using the simplex model: application to repeatedly measured weight in a sample of 164 twins. Acta Genet Med Gemellol (Roma) 1990;39:165–72.
36. Peeters M. Genetic and environmental determination of somatic growth and muscular strength during adolescence: A multivariate and longitudinal analysis. Leuven, Belgium: Katholieke Universiteit Leuven, 1995. PhD Thesis.
37. Thomis M, Maes H, Neale MC, Vlietinck R, Loos R, Beunen G. Genetic determination of growth patterns in stature. Behav Genet 1997;27:608.
38. Little BB, Malina RM, Buschang PH. Sibling similarity in annual growth increments in schoolchildren from a rural community in Oaxaca, Mexico. Ann Hum Biol 1990;17:41–7.
39. Koniarek J. The skeletal development in twins. Materiały i Prace Antropologiczne 1988;108:273–85.
40. Loesch DZ, Huggins R, Rogucka E, Hoang NH, Hopper JL. Genetic correlates of menarcheal age: a multivariate twin study. Ann Hum Biol 1995;22:470–90.
41. Loesch DZ, Hopper JL, Rogucka E, Huggins RM. Timing and genetic rapport between growth in skeletal maturity and height around puberty: similarities and differences between girls and boys. Am J Hum Genet 1995;56:753–9.
42. Towne B, Blangero J, Parks JS, Brown MR, Roche AF, Siervogel RM. Analytic approaches to the study of genetic influences on normal skeletal maturation. In: Gilli G, Schell LM, Benso L, eds. Human growth from conception to maturity. London: Smith-Gordon, 2002:113–23.
43. Thomis M, Maes HH, Peeters M, Loos R, Lysens R, Claessens A, Vanden Eynde B, Vlietinck R, Beunen G. Genetic control of skeletal maturation during growth. Behav Genet 2001;31:471.
44. Towne B, Blangero J, Duren DL, Williams KD, Dyer T, Aivaliotis MJ, Cuttom CR, Lawrence S, Jha B, Subedi J, VandeBerg JL, Williams-Blangero S. A QTL on chromosome 3p influences the tempo of skeletal maturation in Nepali children from early to late childhood. Abstract 1577. Presented at the annual meeting of The American Society of Human Genetics, October 25–29, 2005, Salt Lake City, Utah, USA. Available at: http://www.ashg.org/genetics/ashg05s/. Accessed 6 September 2006.
45. Towne B, Czerwinski SA, Demerath EW, Blangero J, Roche AF, Siervogel RM. Heritability of age at menarche in girls from the Fels Longitudinal Study. Am J Phys Anthropol 2005; 128:210–9.
46. Treloar SA, Martin NG. Age at menarche as a fitness trait: nonadditive genetic variance detected in a large twin sample. Am J Hum Genet 1990;47:137–48.
47. Malina RM, Ryan RC, Bonci CM. Age at menarche in athletes and their mothers and sisters. Ann Hum Biol 1994;21:417–22.
48. Damon A, Damon ST, Reed RB, Valadian I. Age at menarche of mothers and daughters, with a note on accuracy of recall. Hum Biol 1969;41:160–75.
49. Orley J. Analysis of menarche and gynecological welfare of Budapest school girls. In: Eiben OG, ed. Growth and development : Physique. Budapest: Academiai Kiado, 1977:191–4.
50. Kaur DP, Singh R. Parent-adult offspring correlations and heritability of body measurements in a rural Indian population. Ann Hum Biol 1981;8:333–9.
51. Brooks-Gunn J, Warren MP. Mother-daughter differences in menarcheal age in adolescent girls attending national dance company schools and non-dancers. Ann Hum Biol 1988;15:35–44.
52. Koniarek J. The development of secondary sex characters in male twins. Materiały i Prace Antropologiczne 1988;108:239–51.
53. Orczykowska-Swiatkowska Z. Development of secondary sex characteristics in female twins. Materiały i Prace Antropologiczne 1988;108:253–61.
54. Hirschhorn JN, Altshuler D. Once and again—issues surrounding replication in genetic association studies. J Clin Endocrinol Metab 2002;87:4438–41.
55. Willemsen G, Boomsma DI, Beem AL, Vink JM, Slagboom PE, Posthuma D. QTLs for height: results of a full genome scan in Dutch sibling pairs. Eur J Hum Genet 2004;12:820–8.
56. Arinami T, Iijima Y, Yamakawa-Kobayashi K, Ishiguro H, Ohtsuki T, Yanagi H, Shimakura Y, Ishikawa H, Hamaguchi H. Supportive evidence for contribution of the dopamine D2 receptor gene to heritability of stature: linkage and association studies. Ann Hum Genet 1999;63(pt 2):147–51.
57. Hirschhorn JN, Lindgren CM, Daly MJ, Kirby A, Schaffner SF, Burtt NP, Altshuler D, Parker A, Rioux JD, Platko J, Gaudet D, Hudson TJ, Groop LC, Lander ES. Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. Am J Hum Genet 2001;69:106–16.
58. Göring HHH, Duggirala R, MacCluer JW, Kissebah A, Stern MP, Towne B, Williams-Blangero S, Blangero J. Localization of genetic factors influencing human adult height by genome-wide linkage analysis in large pedigree samples. In: Nicoletti I, ed. Human growth in sickness and in health. Florence, Italy: Edizioni Centro Studi Auxologici, 2004:100.
59. Garn SM, Rohmann C. Interaction of nutrition and genetics in the timing of growth and development. Pediatr Clin North Am 1966;13:353–79.
60. Ogata T, Matsuo N. Sex chromosome aberrations and stature: deduction of the principal factors involved in the determination of adult height. Human Genet 1993;91:551–62.
61. Ogata T. The Y-specific growth gene(s): chromosomal localization and its role in growth control. In: Gilli G, Schell LM, Benso L, eds. Human growth from conception to maturity. London: Smith-Gordon, 2002:133–9.
62. Perusse L, Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, Argyropoulos G, Walts B, Snyder EE, Bouchard C. The human obesity gene map: the 2004 update. Obes Res 2005;13:381–490.
63. Arya R, Demerath E, Jenkinson CP, Goring HH, Puppala

S, Farook V, Fowler S, Schneider J, Granato R, Resendez RG, Dyer TD, Cole SA, Almasy L, Comuzzie AG, Siervogel RM, Bradshaw B, DeFronzo RA, MacCluer J, Stern MP, Towne B, Blangero J, Duggirala R. A QTL for birth weight on chromosome 6q identified in two independent family studies. Hum Mol Genet 2006;15:1569–79.
64. Chen W, Li S, Cook NR, Rosner BA, Srinivasan SR, Boerwinkle E, Berenson GS. An autosomal genome scan for loci influencing longitudinal burden of body mass index from childhood to young adulthood in white sibships: The Bogalusa Heart Study. Int J Obes Relat Metab Disord 2004;28:462–9.
65. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 1998; 62:1198–211.
66. Duren DL, Blangero J, Dyer T, Cole SA, Siervogel RM, Towne B. Bivariate linkage of cortical bone thickness and skeletal age to chromosome 3p. Abstract 1466. Presented at the annual meeting of The American Society of Human Genetics, October 25–29, 2005, Salt Lake City, Utah, USA. Available at http://www.ashg.org/genetics/ashg05s/. Accessed 6 September 2006.
67. Parks JS, Brown MR, Siervogel RM, Towne B. Variation in the estrogen receptor gene and the timing of pubertal growth. In: Gilli G, Schell L, Benso L, eds. Human growth from conception to maturity. London: Smith-Gordon, 2002:141–4.
68. Lorentzon M, Lorentzon R, Backstrom T, Nordstrom P. Estrogen receptor gene polymorphism, but not estradiol levels, is related to bone density in healthy adolescent boys: a cross-sectional and longitudinal study. J Clin Endocrinol Metab 1999;84:4597–601.
69. Tao C, Yu T, Garnett S, Briody J, Knight J, Woodhead H, Cowell CT. Vitamin D receptor alleles predict growth and bone density in girls. Arch Dis Child 1998;79:488–94.
70. Suarez F, Rossignol C, Garabedian M. Interactive effect of estradiol and vitamin D receptor gene polymorphisms as a possible determinant of growth in male and female infants. J Clin Endocrinol Metab 1998;83:3563–8.
71. Raivio T, Huhtaniemi I, Anttila R, Siimes MA, Hagenas L, Nilsson C, Pettersson K, Dunkel L. The role of luteinizing hormone-beta gene polymorphism in the onset and progression of puberty in healthy boys. J Clin Endocrinol Metab 1996;81:3278–82.
72. Towne B, Parks JS, Blangero J, Almasy L, Brown MR, Murphy TC, Demerath EW, Roche A, Siervogel RM. Associations between luteinizing hormone-b polymorphisms and skeletal maturation and growth before and during puberty. Am J Hum Genet 1997;61:A214 1238.
73. Lucarini N, Finocchi G, Gloria-Bottini F, Macioce M, Borgiani P, Amante A, Bottini E. A possible genetic component of obesity in childhood. Observations on acid phosphate polymorphism. Experientia 1990;46:90–1.
74. Endo K, Yanagi H, Hirano C, Hamaguchi H, Tsuchiya S, Tomura S. Association of Trp64Arg polymorphism of the beta3-adrenergic receptor gene and no association of Gln223Arg polymorphism of the leptin receptor gene in Japanese schoolchildren with obesity. Int J Obes Relat Metab Disord 2000;24:443–9.
75. Xinli W, Xiaomei T, Meihua P, Song L. Association of a mutation in the beta3-adrenergic receptor gene with obesity and response to dietary intervention in Chinese children. Acta Paediatr 2001;90:1233–7.
76. Tafel J, Branscheid I, Skwarna B, Schlimme M, Morcos M, Algenstaedt P, Hinney A, Hebebrand J, Nawroth P, Hamann A. Variants in the human beta 1-, beta 2-, and beta 3- adrenergic receptor genes are not associated with morbid obesity in children and adolescents. Diabetes Obes Metab 2004;6:452–5.
77. Gelernter-Yaniv L, Feng N, Sebring NG, Hochberg Z, Yanovski JA. Associations between a polymorphism in the 11 beta hydroxysteroid dehydrogenase type I gene and body composition. Int Obes Relat Metab Disord 2003; 27:983–6.
78. Roth H, Korn T, Rosenkranz K, Hinney A, Ziegler A, Kunz J, Siegfried W, Mayer H, Hebebrand J, Grzeschik KH. Transmission disequilibrium and sequence variants at the leptin receptor gene in extremely obese German children and adolescents. Hum Genet 1998;103:540–6.
79. Hinney A, Hohmann S, Geller F, Vogel C, Hess C, Wermter AK, Brokamp B, Goldschmidt H, Siegfried W, Remschmidt H, Schafer H, Gudermann T, Hebebrand J. Melanocortin-4 receptor gene: case-control study and transmission disequilibrium test confirm that functionally relevant mutations are compatible with a major gene effect for extreme obesity. J Clin Endocrinol Metab 2003;88:4258–67.
80. Karvonen MK, Koulu M, Pesonen U, Uusitupa MI, Tammi A, Viikari J, Simell O, Ronnemaa T. Leucine 7 to proline 7 polymorphism in the preproneuropeptide Y is associated with birth weight and serum triglyceride concentration in preschool aged children. J Clin Endocrinol Metab 2000;85:1455–60.
81. Hung CC, Farooqi IS, Ong K, Luan J, Keogh JM, Pembrey M, Yeo GS, Dunger D, Wareham NJ, O'Rahilly S. Contribution of variants in the small heterodimer partner gene to birthweight, adiposity, and insulin levels: mutational analysis and association studies in multiple populations. Diabetes 2003;52:1288–91.
82. Tremblay A, Bouchard L, Bouchard C, Despres JP, Drapeau V, Perusse L. Long-term adiposity changes are related to a glucocorticoid receptor polymorphism in young females. J Clin Endocrinol Metab 2003;88:3141–5.
83. Stavrou I, Zois C, Ioannidis JP, Tsatsoulis A. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. Hum Reprod 2002;17:1101–5.
84. Boot AM, van der Sluis IM, de Muinck Keizer-Schrama SM, van Meurs JB, Krenning EP, Pols HA, Uitterlinden AG. Estrogen receptor alpha gene polymorphisms and bone mineral density in healthy children and young adults. Calcif Tissue Int 2004;74:495–500.
85. Gorai I, Tanaka K, Inada M, Morinaga H, Uchiyama Y, Kikuchi R, Chaki O, Hirahara F. Estrogen-metabolizing gene polymorphisms, but not estrogen receptor-alpha gene polymorphisms, are associated with the onset of menarche in healthy postmenopausal Japanese women. J Clin Endocrinol Metab 2003;88:799–803.
86. Feigelson HS, Coetzee GA, Kolonel LN, Ross RK, Henderson BE. A polymorphism in the CYP17 gene increases the risk of breast cancer. Cancer Res 1997;57:1063–5.
87. Lai J, Vesprini D, Chu W, Jernstrom H, Narod SA. CYP gene polymorphisms and early menarche. Mol Genet Metab 2001;74:449–57.
88. Worda C, Walch K, Sator M, Eppel W, Tempfer CB, Schneeberger C, Huber JC, Hefler LA. The influence of Nos3 polymorphisms on age at menarche and natural

menopause. Maturitas 2004;49:157–62.
89. Kitagawa I, Kitagawa Y, Kawase Y, Nagaya T, Tokudome S. Advanced onset of menarche and higher bone mineral density depending on vitamin D receptor gene polymorphism. Eur J Endocrinol 1998;139:522–7.
90. Comings DE, Gade R, Muhleman D, Peters WR, MacMurray JP. The LEP gene and age of menarche: maternal age as a potential cause of hidden stratification in association studies. Mol Genet Metab 2001;73:204–10.
91. Nanao K, Hasegawa Y. Polymorphisms at the 5' end of the human gonadotropin-releasing hormone receptor gene are not associated with the timing of menarche in Japanese girls. Eur J Endocrinol 2000;143:555–6.
92. Comings DE, Muhleman D, Johnson JP, MacMurray JP. Parent-daughter transmission of the androgen receptor gene as an explanation of the effect of father absence on age of menarche. Child Dev 2002;73:1046–51.
93. Sedlmeyer IL, Pearce CL, Trueman JA, Butler JL, Bersaglieri T, Read AP, Clayton PE, Kolonel LN, Henderson BE, Hirschhorn JN, Palmert MR. Determination of sequence variation and haplotype structure for the gonadotropin-releasing hormone (GnRH) and GnRH receptor genes: investigation of role in pubertal timing. J Clin Endocrinol Metab 2005;90:1091–9.
94. Kadlubar FF, Berkowitz GS, Delongchamp RR, Wang C, Green BL, Tang G, Lamba J, Schuetz E, Wolff MS. The CYP3A4*1B variant is related to the onset of puberty, a known risk factor for the development of breast cancer. Cancer Epidemiol Biomarkers Prev 2003;12:327–31.
95. Goldstein DB, Hirschhorn JN. In genetic control of disease, does 'race' matter? Nat Genet 2004;36:1243–4.
96. Serre D, Paabo S. Evidence for gradients of human genetic diversity within and among continents. Genome Res 2004;14:1679–85.

The International Growth Standard for Children and Adolescents Project: Environmental influences on preadolescent and adolescent growth in weight and height

Stanley J. Ulijaszek

Abstract

This review has two aims. The first is to identify important environmental influences on the growth of children aged 1 to 9 years and of adolescents, defined as those aged 10 to 19 years. The second is to identify possible environmentally based criteria for the selection of individuals and populations for data collection in the development of an international growth reference for these age ranges. There are many common environmental influences on the growth of children between the ages of 1 and 19 years; the examination and description of these forms the main body of this review. Subsequently, environmental factors influencing adolescent growth only are considered. In both cases, possible selection criteria are put forward. The most important inclusion criteria for both preadolescence and adolescence are good nutrition, lack of infection, and socioeconomic status that does not constrain growth. Additionally, low birthweight, catch-up growth, breastfeeding, and early adiposity rebound have impacts on growth and/or body composition into puberty. Exclusion of children born at low birth and/or experiencing catch-up growth could be most realistically operationalized if populations in which secular trends in growth were either completed or minimal were selected. Although an effect of hypoxia on child and adolescent growth, independent of nutrition, is small at most, many high-altitude populations have high prevalances of low birthweight and should be excluded on this basis. Since all populations are exposed to pollutants, contaminants, and toxicants in varying degrees, they cannot be realistically excluded from the sample frame. However, it may be desirable to exclude populations that are habitually exposed to extremely high levels of environmental pollution, including air pollution, and those living in close proximity to toxic waste. It is impossible to exclude populations and individuals on the basis of their exposure to aflatoxin contamination of food. However, exclusion on the basis of low socioeconomic status or poverty may well act as a proxy for this. There are a small number of populations that show extreme patterns of growth in body size and proportion in preadolescence and adolescence, and these should be excluded from the sample frame.

Key words: Aflatoxins, altitude, catch-up growth, growth, infection, low birthweight, nutrition, pollution, puberty, socioeconomic status

Introduction

The evolution of the human growth curve is characterized primarily by an attenuation of childhood, followed by a relatively brief, intense adolescent spurt [1]. The primary selective pressure underlying this evolutionary trend is not certain; however, the extended period of biological immaturity relative to other mammalian species is associated with high environmental sensitivity and growth plasticity [2]. The sensitivity of human growth to the environment is demonstrated both by the processes of stunting and wasting in response to poor nutrition [3] and of catch-up growth during environmental improvements following episodes of stress [4]. An International Growth Reference for Children and Adolescents needs to describe growth patterns associated with maximal health outcomes and to take into account the health risks associated with processes of developmental plasticity [5].

Known environmental factors that influence growth, body size, and body composition of children postnatally include nutrition [6], infection [7, 8], interactions between the two [8–10], psychosocial stress [11], food contaminants [12, 13], pollution [14], and hypoxia [15, 16]. Most of these factors are conditioned by poverty and socioeconomic status [17, 18]. They are also conditioned historically [19], culturally [20], and politically

The author is affiliated with the Institute of Social and Cultural Anthropology, University of Oxford, Oxford, UK

Please direct queries to the author: Stanley J. Ulijaszek, Institute of Social and Cultural Anthropology, University of Oxford, 51 Banbury Road, Oxford OX2 6PF, UK; e-mail: stanley.ulijaszek@anthro.ox.ac.uk.

[1], interacting with each other, but also with individual genotypes in the production of growth, body size, and body composition. **Figure 1** illustrates clusters of variables that influence growth and body composition outcomes and suggests pathways by which they may act. Although the interactions between genetics and environment in the production of growth outcomes are important and are acknowledged in the figure, they are not within the scope of this article.

Diet, nutrition, disease, hypoxia, pollution, contamination, behavioral toxicants, deprivation, and psychosocial stress can be clustered as proximate environmental factors that can influence growth (**fig. 1**, **box 1**). However, they vary in importance according to circumstance and the age and stage in preadolescence and adolescence. Culture, behavior, socioeconomic status, social status, poverty, and political economy can also be clustered as structurally powerful but distal agents (**fig. 1, box 2**) in the production of growth and body composition outcomes, at all ages and stages of childhood and adolescence. This latter cluster is conditioned historically. In the developing world, the risks associated with poverty include low income, low entitlement [21], poor health infrastructures, and environmental hazards. The entitlement of a household is its ability to acquire food through the legal means available in a society [21]; at the extreme, famine occurs not from generalized food shortage, but from inequalities built into mechanisms for distributing and making food available, including social and economic inequalities. In the industrialized world, the risks for impaired child growth that are associated with low socioeconomic status include single parenthood, overcrowding, low disposable income, paternal ill health, dependence on social welfare, and parental abuse of alcohol and drugs [18, 22].

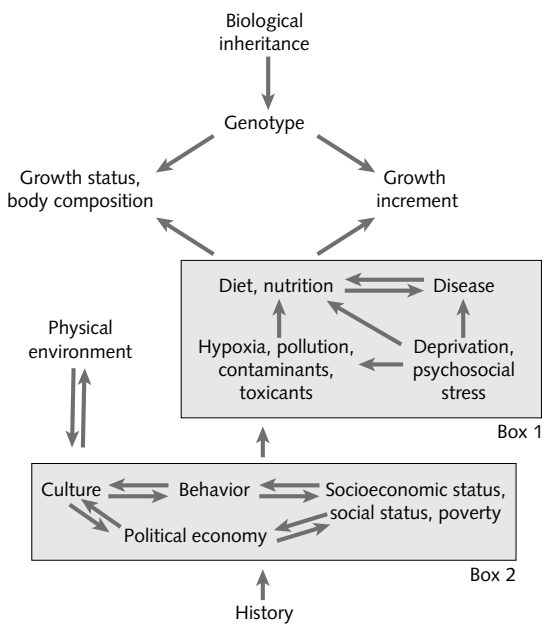

FIG. 1. Factors influencing growth in childhood and adolescence (modified from [2])

In both the developed and the developing worlds, there are similar associations between stature and socioeconomic status, height correlating positively with wealth [18]. Weight, however, does not always relate positively with wealth, overweight and obesity being associated with low socioeconomic status in most industrialized nations [23, 24] and becoming increasingly so among emerging nations undergoing the health transition [25, 26]. Growth stunting in association with overweight was first identified in Peruvian children [27] and more recently in children aged between 3 and 9 years in four nations viewed to be undergoing nutrition transition: Russia, Brazil, South Africa, and China [28]. Childhood obesity is increasing in prevalence across the world [29], with children showing the onset of obesity before 6 years of age often remaining obese into adult life [30, 31]. Critical periods in childhood for the development of obesity include gestation, early infancy, and the period of adiposity rebound between the ages of 5 and 7 years [32].

Early life experiences involving environmental stress, intrauterine growth retardation, poor growth in early childhood, and subsequent catch-up growth can also impact on growth, body composition, and health outcomes later in life [33]. Catch-up growth is an acceleration of child growth rate following either medical or environmental intervention or environmental improvement, such that body size approaches or reaches normality, as defined by appropriate growth references [4]. It can take place at all stages of child growth, including adolescence [34, 35]. However, when the factors responsible for growth faltering or failure are ubiquitous, this process is constrained, and individuals fail to reach their genetic potential for growth and body size [17]. There are possibly three critical periods for the development of obesity and its complications. These include gestation and early infancy, the period of adiposity rebound that occurs between 5 and 7 years of age, and adolescence [32]. The last trimester of pregnancy is critical for appropriate fetal development and birthweight [36], whereas low birthweight followed by subsequent catch-up growth has implications for the subsequent development of obesity in later childhood and adolescence [37–39].

The vast majority of research on environmental influences on human growth has focused on birthweight [40, 41], infancy and infant feeding [42], and early childhood [43]. In contrast, comparatively little attention has been paid to environmental influences on preadolescent growth [44, 45], although Waterlow et al. [3] considered it important to be able to identify poor anthropometric nutritional status among children up to the age of 10 years. Growth and body composition

in adolescence has been researched to a greater extent than in preadolescence, but to a much smaller extent than in children of preschool age [45]. The 1995 World Health Organization monograph "Physical Status: The Use and Interpretation of Anthropometry" [44] pays almost no attention to growth in mid-childhood but devotes one chapter to growth in adolescence. This chapter states that "because growth may be sensitive to nutritional deficit and surfeit, adolescent anthropometry provides indicators of nutritional status and health risk, and may be diagnostic of obesity. The study and understanding of this period of rapid changes are at once, important and difficult" (p. 263). Much of the difficulty stems from the great variation both within and between populations in the timing and magnitude of peak weight and height velocities [46]. Although adolescent growth may be under stronger genetic control than growth in childhood [47, 48], environment can influence both of these measures of adolescent growth and maturation, but to a lesser extent than genetics [1].

The International Growth Reference for Children and Adolescents project seeks to identify environmental factors that can be used as selection criteria for sampling in a large, multinational, multicenter study of the growth of children of school-going age, with the intention of generating growth references suitable for the screening, surveillance, and monitoring of children globally with respect to both undernutrition and obesity. In this respect, the growth patterns of children selected for study should reflect maximal current and future population health. Future health is an important consideration, because environmental factors known to influence growth in early childhood are known to influence growth and body composition in later childhood and adolescence, as well as influencing chronic risk across adult life. This selective review considers environmental influences on growth in stature and weight, since these are the primary measures used in surveillance, screening, and monitoring globally [49]. It begins by considering the environmental factors that can influence growth across both preadolescence and adolescence, and continues by considering specific environmental factors that can influence growth in adolescence. More questions are asked than can be answered at current levels of knowledge, and this review is inevitably incomplete. However, it attempts to inform judgments about appropriate inclusion and exclusion criteria for the sample frame at the population and individual levels.

Overview

There are many problems associated with any definition of how preadolescent and adolescent children should grow. In an examination of the extent to which growth in childhood and puberty varies between populations and shows biological elasticity, Stoltzfus [45] characterized four patterns of population growth that deviate greatly from the normative patterns represented by existing growth references. These include populations that display (1) prepubertal catch-up growth, (2) catch-up growth in puberty, (3) prepubertal stunting combined with catch-up growth in puberty, and (4) prepubertal stunting with no catch-up growth in puberty. Stoltzfus concludes that between-population variation in growth among school-aged children and adolescents is as great as that among children in early childhood. Pattern four is the most common across the developing world [43] and is usually associated with poverty and low socioeconomic status [1, 17]. It is also associated with infant failure to thrive in industrialized nations, where growth-retarded children in infancy have been shown to remain in height and weight deficit at the age of 6 years [49]. Patterns one, two, and three can occur as a result of different types of infant feeding and weaning behavior [42], varying illness management practices [50], dietary manipulation [51], and changing environments during childhood [34, 35]. The latter pattern has been observed to take place across secular trends [52], and during nutrition transition [53].

The term "secular trend" is used to describe marked changes in growth and development of successive generations of human populations living in the same territories [54]. Positive secular trends have been documented among European, European-origin, and Asian populations, where average heights and weights across generations have been shown to be greater, while the adolescent growth spurt has taken place at progressively younger ages [46]. Negative secular trends have also been identified among populations in Africa [55, 56], Papua New Guinea [57], and Central and South America [1]. Since human growth and body size respond with sensitivity to environmental quality, positive secular trends have largely been attributed to improved social, political, nutritional, and health conditions, whereas negative secular trends are often seen as outcomes of environmental, social, or political deterioration [1].

In many places, but not all, the secular trend is associated with the emergence of childhood obesity. Increasing relative weight and fatness with the secular trend in the populations of Western Europe [58–61], Australia [62, 63], Canada [64], the United States [65, 66], and Mexico [67] may reflect some combination of increased availability of dietary energy, increased formula-feeding of infants, and decreased physical activity levels. It may also be due to increased intakes of nonenergy growth-limiting nutrients, such as calcium or zinc, which can influence the positive secular trend independently of energy intake [35]. Hauspie et al. [68] identified more marked secular changes to have taken place in the lower height centiles in many populations,

an observation that may have direct implications for the definition of short stature in a population, and a consideration for the development of growth references for preadolescents and adolescents.

Environmental factors influencing growth patterns in preadolescence and adolescence

The impacts of poverty and low socioeconomic status on growth and malnutrition in preadolescence and adolescence in both the developing and the developed world are well known [17, 69, 70]. Of the various environmental factors influencing the growth of children in developing countries, diet, nutrition, and infection are particularly powerful across most of childhood, but especially in early childhood. In addition to these factors, child neglect and abuse, exposure to industrial pollutants, food contaminants, behavioral toxicants, single parenthood, overcrowding, and parental ill health are often important contributors to growth outcomes, according to circumstance [22]. The nature of interrelationships between proximate (**fig. 1**, **box 1**) and distal (**fig. 1**, **box 2**) influences on child growth has been examined at the macro level in two meta-analyses. In an analysis of the prevalence of risk factors and income poverty for 11 medium- and low income world subregions using World Health Organization and World Bank data, Blakely et al. [70] identified strong relationships between poverty and childhood malnutrition, access to unsafe water and sanitation, and exposure to indoor air pollution. Using different sources, a global cross-sectional survey of the determinants of stunting in children below the age of 5 years has shown the most important factors to be dietary energy availability, female literacy, and gross national product [71].

Such relationships change across the course of the nutrition transition [72]. For example, longitudinal analysis of panel data from 63 developing countries suggests that increased national food availability across the period from 1970 to 1996 has had a strong positive impact on child nutritional status, with per capita increase in food supply having the strongest effects in sub-Saharan Africa and South Asia but only weak or nonexistent effects in countries with higher food security in 1970, particularly in the Near East and North Africa [73]. There is also the shift in the burden of poor diets, inactivity, and obesity from the rich to the poor [72].

In the developed world, relationships between poverty and child growth persist, but perhaps to a lesser extent than prior to positive secular trends that took place from the late nineteenth and across the twentieth century [46]. However, poverty in developed nations may have less environmental bearing on child growth but more on obesity than in developing nations. For example, examination of data from the US National Health and Nutrition Examination Survey has shown poverty to be predictive of poor nutrition among preschool children but not among school-aged children [74].

Nutrition

Both dietary quantity and quality can influence child growth and body composition, as can infant-feeding patterns. Although there appears to be no difference in growth status between breastfed and formula-fed children by the time they reach school age, breastfed infants are less likely to become obese [42]. Deficiencies of energy [75], protein [76, 77], and zinc [75, 78] have been implicated in growth faltering, while diets high in fat are associated with weight gain [79] and obesity [80]. Vegetarian and vegan diets are associated with zinc deficiency [81], while the primarily vegetarian diets of schoolchildren in rural Egypt, Kenya, and Mexico have been shown to be deficient in vitamin A, vitamin B_{12}, riboflavin, calcium, iron, and zinc [82]. In a study from the Netherlands, children aged 0 to 10 years consuming macrobiotic diets with adequate protein and energy intakes and protected from bacterial contamination were shown to have a growth pattern similar to that of children from poor developing countries [83]. Furthermore, various micronutrient intervention studies have demonstrated the importance of specific deficiencies in human growth. For example, multimicronutrient interventions have been shown to have a positive effect on growth in both weight and height of children in India [84], while combined vitamin A and zinc supplementation has been shown to improve impaired intestinal barrier function and linear growth shortfalls in young children in Brazil [78]. In a review of published, randomized clinical trials that examined the impact of administration of micronutrients to infants and children on linear growth, Bhandari et al. [85] concluded that zinc and iron have a modest effect on linear growth in deficient populations, whereas vitamin A is unlikely to have any important effect. Partial support for this position comes from subsequent meta-analyses of randomized, controlled intervention trials conducted to assess the effects of micronutrient interventions on the growth of children below 18 years of age [86], which have concluded that deficiencies of vitamin A and iron only cause growth faltering when they are severe.

Diets low in fat have been shown to affect child growth. For example, diets containing less than 30% of energy as fat have adverse effects on child growth beyond infancy, but it is unclear whether this is due to low total energy intake or to an associated effect on dietary energy density and/or other nutrients [87]. Children with low fat intake are at risk for inadequate intakes of fat-soluble vitamins [88] such as vitamin A, although this may only be important for child growth in populations with extremely low fat and vitamin A

intakes. Children with very low fat intakes may be at risk for inadequate intakes of other nutrients too. For example, in the United States, 10-year-olds consuming less than 30% of their dietary energy as fat are less likely to meet the recommended dietary allowances for vitamin B_6, vitamin B_{12}, vitamin E, thiamine, riboflavin, and niacin [89].

This relationship is likely to be mediated by socioeconomic status. For example, self-selected low fat intake by children aged 8 to 12 years from average to high socioeconomic backgrounds in Norway does not compromise their intake of either macro- or micronutrients or dietary energy [90]. However, children of low socioeconomic status who consume diets with below average fat are likely to consume proportionately more sugar and refined carbohydrates [91], both of which are more affordable than lean meats, fish, fresh vegetables, and fruit [92, 93], thus limiting their intakes of micronutrients. Children of higher socioeconomic status consuming low-fat diets are more likely to have greater access to foods that are richer in micronutrients than energy-dense foods.

Cultural manipulation of diet can override political-economic influences on diet quality, when individuals, groups, and populations consume vegetarian or vegan diets for religious or ideological reasons. For example, Dutch children brought up on macrobiotic diets are deficient in dietary energy from fat, calcium, riboflavin, and vitamin B_{12} [94], experiencing growth faltering after the time of weaning but showing catch-up growth by the age of 9 years [83]. The heavy use of staple foods that carry strong symbolic meaning and cultural identity for societies can also limit intake of energy and micronutrients. For example, in Papua New Guinea, consumption of staples rich in symbolic meaning as well as carbohydrates has been shown to be associated with poor growth of young children [95]. In Brazil, obese children have been shown to have low zinc intake and poor zinc nutritional status [96]. This may be one way in which poor diet quality may contribute to the emergence of obesity in association with short stature, which was first identified in Peruvian children in the 1980s [27] and subsequently in children in Russia, Brazil, South Africa, and China [28].

Infection

Infectious disease is an important cause of growth retardation in poor communities [9]. Among the diseases and disease categories associated with either growth faltering or poor weight gain are diarrhea [97–100], respiratory tract infections [97], intestinal parasitic infections [101], malaria [100, 102], HIV infection [100, 103–106], schistosomiasis [107, 108], *Cryptosporidium parvum* infection [109, 110] and *Helicobacter pylori* infection [111]. All of these are associated with poor nutritional status [69, 112–114]. Deficiencies of vitamin A and zinc are strongly associated with infectious disease risk [114]. Infections that influence nutritional status and linear growth are either acute and invasive, provoking a systemic response (such as dysentery and pneumonia), or chronic, affecting the host over a sustained period (including gut helminth infections). Infections can diminish linear growth by affecting nutritional status [115] by way of decreased food intake, impaired nutrient absorption, direct nutrient losses, increased metabolic requirements, catabolic losses of nutrients, and/or impaired transport of nutrients to target tissues [114]. In addition, induction of the acute-phase response and host elaboration of proinflammatory cytokines [102] may contribute to growth faltering, because they directly inhibit the process of bone remodeling that is needed for long bone growth [7].

Infections are associated with poverty. Farmer [116] has described the extensive ways in which poverty and social inequality are embodied as differential risks for infection with HIV and tuberculosis in developing countries, and Walls and Shingadia [117] identified overcrowding, poverty, and the HIV epidemic as contributing to the resurgence of tuberculosis globally. Bates et al. [118] identified poverty as a key factor operating at the individual, household, and community levels in increasing vulnerability to malaria, tuberculosis, and HIV infection. Such relationships are not exclusive to developing countries. For example, in Scotland, *Helicobacter pylori* infection has been found to be strongly associated with poverty among children of all ages [119].

Poverty also underpins the nutrition and infection interactions that affect child growth. Tuberculosis, associated with household crowding and poverty, may not be directly associated with growth faltering, but it is associated with nutritional status. Children born to HIV-infected women who are vitamin A deficient during pregnancy are more likely to experience growth failure [120]. Furthermore, HIV infection, malaria, and diarrheal disease adversely affect the growth of preschool-aged children and are associated with an increased prevalence of vitamin A deficiency [100].

Environmental pollutants, food contaminants, and behavioral toxicants

Pollution and contamination are globally widespread, and it is unlikely that there is any human population that is completely unaffected by them [121]. Of the industrial contaminants that affect postnatal growth, lead and polychlorinated biphenyls (PCBs) are the most important [121]. Food contamination by aflatoxins, which can influence health and growth, is an outcome of agricultural practice and of food storage and processing practices. Behavioral toxicants, exposure to which comes from the use of alcohol, cigarettes, and narcotics, can affect birthweight, and they merit scrutiny in relation to child growth.

Levels of pollutants that human populations are exposed to vary markedly, depending partly on the degree of, and proximity to, industrialization [121]. Generalized air pollution has been identified as an environmental risk factor for poor growth of children in communities in Silesia and Belgium [121]. Pollution from hazardous waste sites may also impair growth and development. For example, children born and raised in Love Canal, New York, where a leaking hazardous waste site was discovered in 1978, were significantly shorter than children from a community similar in social characteristics but located far from the site [122].

With respect to specific components of pollution, preliminary evidence suggests that low levels of lead (below 25 μg/dL), which are now commonplace in industrialized and industrializing countries, can affect postnatal physical growth and development [123]. Evidence of an effect of contamination by PCBs, polychlorinated dibenzofurans, and polychlorinated quaterphenyls on child growth comes from an epidemiologic study in western Japan, where consumption of contaminated rice oil in 1979 [124] led to depressed growth in postnatally exposed children 4 years after exposure [125].

The risk of environmental pollutant exposure to child growth is increasing as developing countries industrialize. For example, China has seen pronounced increase in anthropogenic lead during the past two decades; this has been attributed to rapid economic growth and the lack of waste-control practices [126]. In Taiwan, drinking water and factories in the neighboring areas are significant sources of high blood lead, in addition to occupational lead exposure [127].

Aflatoxins are mold metabolites produced by toxigenic strains of *Aspergillus* species, a number of which are hepatotoxic and immunotoxic. They have been associated with child growth faltering in Benin [12, 13] and Togo [12]. Primary commodities susceptible to aflatoxin contamination include corn, peanuts, cottonseed, and animal-derived foods such as milk when the animal is fed aflatoxin-contaminated feed [128]. Significant dietary contamination by aflatoxins has been demonstrated in Benin, West Africa [129], Southwestern Nigeria [130, 131], Iran [132], India [133], the Philippines, Thailand, and Indonesia [134], as well as Brazil, Argentina, Uruguay, Paraguay, and Venezuela [135]. Excessive aflatoxin contamination is not global, however. Other studies in Brazil [136] and the Philippines [137] showed low levels of aflatoxin food contamination, whereas in Taiwan [138], in Australia [139], and across the European Union [140], the levels are low. Although aflatoxin monitoring and control in the United States is as rigorous as in the United Kingdom and Australia [139], the United States has experienced aflatoxin contamination of corn for both human and animal consumption during drought years [141].

Risks associated with aflatoxin-contaminated foods can be reduced through the use of multiple processing and decontamination procedures, including physical cleaning and separation procedures [128]. Although aflatoxins are partially destroyed during nixtamalization, the alkaline cooking procedure employed to prepare tortillas in Mexico [142], high levels of aflatoxin often remain in the food [143]. However, the governments of Brazil, Argentina, Colombia, Venezuela, and Uruguay, the main grain-exporting countries in South America, have established food-safety guidelines and regulations for aflatoxin control in national food supplies [135]. With increasing global trade, countries of the developing world wishing to export food commodities increasingly will have to comply with rigorous standards of food safety, including aflatoxin contamination. However, although aflatoxin-low food products may be approved for sale internationally, there is a risk that the same high standards for aflatoxin safety might not be universally applied across developing countries.

Prenatal exposures to cocaine, alcohol, and cigarettes are linked to decreased birthweight and length. However, studies attempting to identify whether such growth deficits persist into childhood have led to differing conclusions [144]. Since studies vary in the number of subjects, cohort characteristics, measurement of exposure, and control for potential confounders, Nordstrom-Klee et al. [144] highlight the need for large-scale, well-designed studies to clearly examine the separate contributions of varying prenatal exposures and the effect of the magnitude and timing of these exposures on childhood growth deficits.

High altitude

Relative to lowland populations, members of populations living at high altitude are generally born with low birthweight, whether in Colorado [145] or Bolivia [146], and experience slow and prolonged postnatal growth [147, 148], with delayed puberty and smaller adult body size [149]. For example, Han children born and raised at high altitudes in China are smaller and lighter into adolescence than those born and raised at low altitudes. Slower growth at high altitudes is a consequence of hypoxia [147], poor economic conditions, and nutritional inadequacy [149]. A study of European children of higher socioeconomic status migrating to high altitudes in the Andes has shown them to be slightly shorter and lighter than their peers of same socioeconomic status living at sea level, indicating that high-altitude hypoxia has a small but independent effect on growth [150, 151].

Life history and critical periods of development

Stoltzfus [45] characterized four patterns of population growth that deviate greatly from the normative patterns represented by existing growth references. The first three of these, representing prepubertal catch-up

growth, catch-up growth in puberty, and prepubertal stunting combined with catch-up growth in puberty, respectively, are associated with critical periods of development that can have long-term implications for health and body composition in later childhood [36, 152]. Body fatness reaches a postinfancy low level typically between the ages of 5 and 7 years, followed by increased body fatness, a phenomenon termed adiposity rebound by Rolland-Cachera et al. [31]. Early adiposity rebound has been associated with earlier age at menarche [152, 153] and increased relative weight and obesity later in life, including during adolescence [31, 154–156]. Early growth restriction followed by catch-up growth is also associated with the development of abdominal obesity [38], whereas higher growth velocity in early childhood, prior to adiposity rebound, has been shown to be associated with greater fatness and obesity in subsequent years in Brazil [39]. In the United States, low levels of vigorous physical activity and high levels of television viewing have been associated with fatness in children during the adiposity rebound period [154]. The combination of small size at birth and during infancy, followed by accelerated weight gain from the ages of 3 to 11 years, predicts large differences in the cumulative incidence of coronary heart disease, non-insulin-dependent diabetes, and hypertension in later life [152]. Therefore, new international growth references should be able to identify catch-up growth across preadolescence and adolescence, as well as adiposity rebound.

Growth and healthy practice

In contrast to negative impacts on growth, there is evidence that healthy practices can influence growth in a positive way, at least in young children. In rural India, studies on the development of eating behavior suggest that mothers of young children make food choices that fit into their budgets, but that they are also influenced by new information [157]. Their food choices are largely conditioned by traditional beliefs, some of which have positive effects on nutrition and on the growth and psychosocial development of infants and children of preschool age. In China, breastfeeding until 12 months of age and mothers' knowledge and practice of good nutrition are associated with positive growth and nutritional status of their infants [158]. However, school attendance or employment can take older children away from the dietary influence of the household for significant periods of time. Both income and institutional feeding arrangements may make such children differ from younger ones in their dietary habits and nutritional status. Although many countries have implemented school-feeding programs, only one published source could be identified in which the impact of such intervention on growth was evaluated. A randomized, controlled trial in which breakfast was served to children in grades 2 to 5 of 16 rural Jamaican schools for one year resulted in only small benefits to children's anthropometric nutritional status [159].

Body proportion

Extreme patterns of body proportionality development have been found in a small number of global minority populations. These include Australian aboriginal people, East African pastoralists, and African pygmies. Among pastoralist Karimojong and Turkana children, linearity of physique is related to narrow skeletal breadth, a trait that emerges early in childhood and appears to have a substantial genetic component [160, 161]. Little [162] showed that three African pastoralist groups have patterns of linear growth in childhood and adolescence that are only slightly delayed relative to African-American reference values, but with much lower body weight at any given height. Leg length and fat distribution of Australian Aboriginal people differ from those of non-Aboriginal populations in childhood, adolescence, and adulthood [163–165]. Different body proportions among Aboriginal people lower the range of normal body-mass index (BMI) values for them [166]; the World Health Organization (WHO) suggested that in Aboriginal adults, this normal range may lie between 17 and 22 kg/m^2 [29] rather than the 19 to 25 range recommended for European populations. Preliminary evidence suggests that the same should apply to Aboriginal children [167]. African pygmies of known age from birth to 5 years show progressive growth faltering relative to United States growth references from birth onward [168]. They also do not have a period of accelerated adolescent growth, making them among the very shortest of all human populations in adulthood [169]. This shortness is due to resistance to the growth-promoting actions of both insulin-like growth factor I (IGF-I) and growth hormone [170]; the IGF-I receptor has been identified as the locus governing their short stature [171].

Psychological stress, deprivation, and neglect

The idea that psychological stress causes growth faltering in some children was first put forward by Widdowson [172], who published evidence that the presence of a sadistic schoolteacher caused child growth in an orphanage to falter, despite a concurrent increase in the amount of food eaten. Furthermore, family conflict was shown to be associated with short stature in childhood as well as short adult height in the British 1958 cohort study [173].

An association between social deprivation and impaired physical growth comes from studies of children classified as having psychosocial dwarfism [174], psychosocial short stature [175], or hyperphagic short stature [176] in industrialized countries. The majority of such children have been characterized as having below-average birthweight, having backgrounds in which social or emotional problems are common, and

exhibiting abnormal eating patterns and/or behavior problems [177], as well as having growth hormone insufficiency, which is reversible with a change in social environment [174]. Many of the symptoms characteristic of psychosocial dwarfism are nonspecific stress responses, whereas hypothalamic pathologies are viewed as accounting most of the clinical features of hyperphagic short stature [176]. However, the dominant mechanism is viewed to be nutritional, by way of appetite loss and anorexia [178]. Deprivation and stress are not synonymous, although they often co-occur, especially among families of low socioeconomic status.

General selection criteria

Selection criteria for the sample of children and adolescents should acknowledge the importance of environmental effects on growth and body composition in utero and in the preschool years and the influence such effects may have on growth and body composition in subsequent childhood and adolescence. Thus, children born with low birthweight should be excluded. The World Health Organization has accepted a definition of low birthweight as being below 2,500 grams, the measurement being taken preferably within the first hours of life, before significant postnatal weight loss has occurred [179]. The selection of populations in which secular trends in growth were either completed or minimal would minimize the proportion of children in the sample born with low birthweight.

Given the importance of catch-up growth for subsequent growth and body composition in later childhood and adolescence, it may be useful to exclude individual children who have experienced catch-up growth in earlier life. Exclusion of children born with low birthweight might be one simple way of achieving this, at least with respect to catch-up growth in infancy. Identifying children experiencing catch-up growth postinfancy is likely to be very difficult among some populations, and it might be better to focus on population-level selection criteria, which would either include groups whose growth patterns were stable or exclude populations undergoing the nutrition transition. It would be useful if such populations were also politically and economically stable, so that the extent to which infants and children experienced the combination of low birthweight and catch-up growth was small, as was the proportion of individuals experiencing early adiposity rebound. In this way, all known patterns of growth that deviate from reference norms and that can be associated with long-term implications for health and body composition in later childhood could be excluded. Individuals selected for inclusion should also have been breastfed, because breastfeeding appears to minimize the risk of developing obesity after 5 years of age [42].

Another selection criterion should be that children for the study should come from populations in which socioeconomic status does not constrain child growth, and perhaps also where food balance sheets indicate a daily per capita dietary energy supply between limits associated with minimal levels of undernutrition and overnutrition at national level. Defining such limits is beyond the scope of this article but would not be difficult to model. Populations consuming vegetarian diets low in fat, as determined from food balance records, could also be considered for exclusion. It may be valuable to exclude vegans from the study sample, not only because they show growth faltering in early life, but also because they may experience subsequent catch-up growth. Given the lack of rigorous evidence in support of a positive effect of school-meal programs on child growth, this cannot be a selection criterion for this project.

A number of infectious diseases are known to be directly associated with growth faltering and/or low weight gain, including diarrhea, intestinal parasitic infections, malaria, schistosomiasis, and infections by HIV, *Cryptosporidium parvum*, and *Helicobacter pylori*. There is a greater range of infections, acute, invasive, and chronic, that are associated with nutritional status. Most of these infections are associated with poverty; employing socioeconomic status that does not constrain growth as a selection criterion would remove most of these effects from the study sample. The additional exclusion of populations with high morbidity and mortality from infectious disease, regardless of socioeconomic status, would remove most of the remaining effects of infectious disease. Seasonality of growth occurs among populations of both developed and developing countries. In the former, the effects are quite subtle climatic ones, while in the latter, they are largely due to seasonal variation in food availability and exposure to infectious disease [180]. Exclusion of populations on the basis of poor nutrition and high morbidity and mortality from infectious disease would remove the most important seasonal effects on growth.

Postnatal growth of populations at high altitudes is slow and prolonged, with delayed puberty and smaller adult body size relative to lowland populations and Western norms. This effect is to varying degrees independent of socioeconomic status and nutritional state. Low birthweight is also very common at high altitudes. The International Growth Reference for Children and Adolescents project should exclude populations living at high altitudes because of the usually high prevalence of low birthweight and the associated environmental constraints of undernutrition, poverty, and low socioeconomic status found among many of them.

It is unrealistic to exclude all populations that are exposed to pollutants, contaminants, and toxicants, since these appear to be ubiquitous. However, at very

high doses, general pollution and specific exposure to PCBs can affect child growth, whereas exposure to lead can affect growth at moderate to low levels. Thus it may be desirable to exclude populations that are habitually exposed to extremely high levels of environmental pollution, including air pollution, and those living in close proximity to toxic waste. Exposure to lead is associated with poverty [121], and exclusion on the basis of poverty or low socioeconomic status should also exclude the majority of subjects who are exposed to lead. Use of lead-glazed ceramics has been shown to be associated with significant elevations in blood lead levels among children in Mexico City [181]. Therefore it may be also prudent to exclude populations habitually exposed to high levels of lead from indigenous pottery that is characteristically underfired and that leaches lead from the glaze into foods and drink. Agricultural contaminants such as aflatoxins appear to be widely distributed. It is likely that economic development and globalization of food trade, accompanied by food-contamination legislation, will result in increased monitoring of food contamination and eventual declines in aflatoxin levels in human food supplies across the developing world. Thus, although it is impossible to exclude populations and individuals on the basis of their exposure to aflatoxin contamination, exclusion on the basis of low socioeconomic status or poverty may well act as a proxy, at least into the near future. It is not unequivocally clear whether postnatal exposures to cocaine, alcohol, and cigarettes are linked to poor growth and body composition outcomes. Although these cannot be used as exclusion criteria, given present knowledge, they do not have positive influences on child growth.

There are a small number of populations that show extreme patterns of growth in body proportion in preadolescence and adolescence, or of body size across puberty. Selection criteria should perhaps exclude populations that show extreme body proportionality emergent across the course of growth and development, such as East African pastoralists, Australian Aboriginal people, and African pygmies, from the sample frame.

Environmental factors influencing variability in adolescent growth

In this section, environmental influences that are specific to adolescent growth are considered. Although broad descriptive studies of adolescent growth are plentiful, there are few that have critically evaluated the relative importance of specific environmental factors in this process. This is undoubtedly because of the great variation both within and between populations in the timing and magnitude of peak weight and height velocities. In general, adolescent growth is seen to be sensitive to nutritional deficit and surfeit [44, 182]. Although this might be the dominant environmental influence at this stage of growth, this section considers the slim literature concerning other factors as well. Environmental factors that can influence growth in adolescence as well as preadolescence are considered in the previous section, and are not discussed here. These include micronutrient deficiencies, tuberculosis and *Helicobacter pylori* infection, generalized air pollution, exposure to hazardous waste, life history and critical periods of development, and psychological stress and deprivation.

For many industrialized nations, there have been secular trends in the timing and size of the pubertal growth spurt that have been taken as evidence for nutritional improvement [182]. In addition, many of the socioeconomic differences between groups in growth in adolescence have been attributed to nutrition [182].

There appears to be only one longitudinal study that demonstrates direct nutritional effects on growth in adolescence [183], which was conducted among girls in Boston, USA. Those who consumed more dietary energy and animal protein than average 2 years before peak growth were shown to have higher peak height velocity (PHV) than average [183]. As for possible positive impacts of school-meals programs on adolescent growth, there is no rigorous evidence at present that they have any influence on growth or nutritional status in puberty. In a study of Nigerian adolescents aged between 13 and 18 years living in boarding houses, the pupils often missed school meals, and their nutritional status was independent of the amount of institutional feeding accepted by them [184].

The impact of infection is considered to be of much lesser importance than that of nutrition, perhaps because the immune system has matured and adaptive immunity is largely in place by adolescence [185], and experience of most common infections of early childhood is much reduced by this stage of growth. However, perinatal HIV-1 infection has been shown to interfere with sexual maturation in children surviving this infection into adolescence in Italy [104] and the United States [105]. Furthermore, French children suffering from a number of chronic diseases have been shown to experience delayed onset of puberty and a reduced pubertal growth spurt [115]. Two interactions between nutrition and infection have been demonstrated to have specific impacts on adolescent growth and relative weight. *Helicobacter pylori* infection and iron-deficiency anemia have been shown to be associated with delayed pubertal growth [186], whereas malarial infection in early puberty is associated with high parasite density and low relative weight [102].

Exposure to environmental pollutants in adolescence might be expected to be different from exposure in earlier life, since the likelihood of occupational exposure increases. Perinatal exposure to polybrominated biphenyl (PBB) has been shown to be associated with earlier age at menarche [187], while such early exposure to

PCB is associated with delayed sexual maturation in both males and females [188]. In the one study of the concurrent effects of most common pollutants to which children might be exposed, Denham et al. [189] found attainment of menarche to be sensitive to relatively low levels of lead and certain PCB congeners. The relationship between age at menarche and PCB congeners observed by Denham et al. [189] is concordant with the observations of Blanck et al. [187] in relation to age at menarche and PBB, probably because of similarities in structure of the toxicants themselves. Dichlorodiphenyltrichloroethane (DDT) is a chemical that was once used widely in agriculture but is now limited largely to public health use, especially in malarial vector control programs in nations where equally effective and affordable alternatives are not locally available [190]. The one study in the United States that rigorously examined the possible effects of prenatal DDT exposure on pubertal growth and development found no effect [191]. In the developed countries, the teenage years are known for experimentation and use of behavioral toxicants. For example, In Australia, cigarette smoking and the use of intoxicants and narcotics by adolescents has been shown to increase with pubertal stage, independent of age [192]. Although maternal smoking, alcohol, and drug use can affect the growth status of the fetus and infant [193], as can passive smoking [194, 195], it is not clear whether the uptake of behavioral toxicants by adolescents has any effect on their growth [196].

There is delayed sexual maturation among many, but not all, high-altitude populations relative to lower-altitude populations [147, 197], as well as a late and poorly defined adolescent growth spurt, at least among Andean populations [198]. However, the importance of hypoxia to delayed skeletal and sexual maturation relative to nutritional stress is low. Two examples can illustrate this. Ethiopians living at high altitude have better nutrition, growth, and socioeconomic conditions than low-altitude populations, and this is reflected in their better growth rates [199]. The second is a study of a well-nourished European-origin population living at high altitude. The timing and size of the pubertal growth spurt of children taking part in the Denver longitudinal growth study [200] were similar to that of the majority US population. In the latter case, hypoxic stress at an altitude of 1,600 m is too small to affect pubertal growth.

Known impacts of psychological stress on pubertal growth are limited sexual and skeletal maturity. A prediction based in evolutionary theory is that environmental instability should favor early reproduction [201]. This is borne out by studies of girls exposed to familial distress or living in dysfunctional households, who have been shown to have an earlier age of menarche [202, 203] and to be taller at early and middle stages of puberty [204] than girls who were not habitually exposed to such stress.

Additional selection criteria for adolescents

Most of the environmental factors associated with growth in adolescence are common to growth in preadolescence, the most important of which are nutrition, infection, and the interactions between the two. Additionally, birthweight, catch-up growth, breastfeeding, and early adiposity rebound have impacts on growth and/or body composition into puberty. As with preadolescent growth, the International Growth Reference for Children and Adolescents project should exclude children exposed to poor nutrition, a significant infectious disease burden, and low socioeconomic status. Since the primary environmental insult causing the late and poorly defined adolescent growth spurt at high altitudes appears to be nutritional, there may be no additional gain to the project by excluding such populations.

Evidence is lacking to suggest that use of behavioral toxicants influences growth in adolescence, although adolescents exposed to extremely high levels of air pollution and/or toxic waste should be excluded. As with preadolescent children, it is unrealistic to exclude all adolescents that are exposed to lower levels of pollution, although exposure to PCBs and moderate levels of lead can affect puberty, but excluding those exposed to very high levels is probably prudent. There is no knowledge of the possible impacts of aflatoxin exposure on growth in adolescence, and no recommendation can be made concerning these substances as selection criteria. As with preadolescent growth, selection criteria should exclude groups that show major differences in body proportion in adolescence, such as East African pastoralists, Australian aboriginal people, and African pygmies.

References

1. Bogin B. Patterns of human growth. Cambridge, UK: Cambridge University Press, 1999.
2. Johnston FE. The ecology of post-natal growth. In: Ulijaszek SJ, Johnston FE, Preece MA, eds. Cambridge encyclopedia of human growth and development. Cambridge, UK: Cambridge University Press, 1998:315–9.
3. Waterlow JC, Buzina R, Keller W, Lane JM, Nichaman MZ, Tanner JM. The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. Bull World Health Organ 1977;55:489–98.
4. Prader A, Tanner JM, von Harnack GA. Catch-up

growth following illness or starvation. An example of developmental canalization in man. J Pediatr 1963; 62:646–59.
5. Ulijaszek SJ. Long-term consequences of early environments on human growth: a developmental perspective. In: Henry CJK, Ulijaszek SJ, eds. Long-term consequences of early environment. Growth, development and the lifespan developmental perspective. Cambridge, UK: Cambridge University Press, 1996:25–43.
6. Allen LH. Nutritional influences on linear growth: a general review. Eur J Clin Nutr 1994;48:S75–89.
7. Stephensen CB. Burden of infection on growth failure. J Nutr 1999;129(2S suppl):534S–8S.
8. Bhan MK, Bahl R, Bhandari N. Infection: How important are its effects on child nutrition? In: Martorell R, Haschke F, eds. Nutrition and growth. Philadelphia, Pa, USA: Lippincott Williams and Wilkins, 2001:197–217.
9. Scrimshaw NS, Taylor CI, Gordon JE. Interactions of nutrition and infection. Geneva: World Health Organization, 1968.
10. Ruel MT. The natural history of growth failure: importance of intrauterine and postnatal periods. In: Martorell R, Haschke F, eds. Nutrition and growth. Philadelphia, Pa, USA: Lippincott Williams and Wilkins, 2001:123–53.
11. Powell GF, Brasel JA, Raiti S, Blizzard RM. Emotional deprivation and growth retardation simulating idiopathic hypopituitarism: I Clinical evaluation of the syndrome. N Engl J Med 1967;276:1271–8.
12. Gong YY, Egal S, Hounsa A, Turner PC, Hall AJ, Cardwell KF, Wild CP. Determinants of aflatoxin exposure in young children from Benin and Togo, West Africa: the critical role of weaning. Int J Epidemiol 2003;32:556–62.
13. Gong YY, Hounsa A, Egal S, Turner PC, Sutcliffe AE, Hall AJ, Cardwell KF, Wild CP. Postweaning exposure to aflatoxin results in impaired child growth: a longitudinal study in Benin, West Africa. Environ Health Perspect 2004;112:1334–8.
14. Schell LM. Environmental toxicants. In: Ulijaszek SJ, Johnston FE, Preece MA, eds. Cambridge encyclopaedia of human growth and development. Cambridge, UK: Cambridge University Press, 1998:343–5.
15. Frisancho AR. Developmental adaptation to high altitude hypoxia. Int J Biometeorol 1977;21:135–46.
16. Mueller WH, Schull VN, Schull WJ, Soto P, Rothhammer F. A multinational Andean genetic and health program: growth and development in an hypoxic environment. Ann Hum Biol 1978;5:329–52.
17. Martorell R, Mendoza F, Castillo R. Poverty and stature in children. In: Waterlow JC, ed. Linear growth retardation in less developed countries. New York: Raven Press, 1988:57–70.
18. Bogin B. Social and economic class. In: Ulijaszek SJ, Johnston FE, Preece MA, eds. Cambridge encyclopaedia of human growth and development. Cambridge, UK: Cambridge University Press, 1998:399–401.
19. Fogel RW. Physical growth as a measure of the economic well-being of populations: the eighteenth and nineteenth centuries. In: Falkner F, Tanner JM, eds. Human growth: a comprehensive treatise. Vol 3. New York: Plenum Press, 1986.
20. Ulijaszek SJ. Human adaptation and adaptability. In: Ulijaszek SJ, Huss-Ashmore RA, eds. Human adaptability. Past, present and future. Oxford, UK: Oxford University Press, 1997:7–16.
21. Sen A. Poverty and famines: an essay on entitlement and deprivation. Oxford, UK: Clarendon Press, 1981.
22. Schell LM. Risk focusing: an example of biocultural interaction. In: Huss-Ashmore R, Schall J, Hediger M, eds. Health and lifestyle change. MASCA Research Papers in Science and Archaeology 1991;9:137–44.
23. Sobal J. Obesity and socioeconomic status: a framework for examining relationships between physical and social variables. Med Anthropol 1991;13:231–47.
24. de Garine I, Pollock N. Social aspects of obesity. New York: Gordon and Breach, 1995.
25. Peña M, Bacallao J. Obesity among the poor: an emerging problem in Latin America and the Caribbean. In: Obesity and poverty: A new public health challenge. Peña M, Bacallao J, eds. Washington, DC: Pan American Health Organization, 2002.
26. Xu F, Yin XM, Zhang M, Leslie E, Ware R, Owen N. Family average income and body mass index above the healthy weight range among urban and rural residents in regional Mainland China. Public Health Nutr 2005;8:47–51.
27. Trowbridge FL, Marks JS, Lopez de Romana G, Madrid S, Boutton TW, Klein PD. Body composition of Peruvian children with short stature and high weight-for-height. II. Implications for the interpretation of weight-for-height as an indicator of nutritional status. Am J Clin Nutr 1987;46:411–8.
28. Popkin BM, Richards MK, Montiero CA. Stunting is associated with overweight in children of four nations that are undergoing the nutrition transition. J Nutr 1996;126:3009–16.
29. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:i–xii, 1–253.
30. Poskitt EM, Cole TJ. Do fat babies stay fat? Br Med J 1977;1:7–9.
31. Rolland-Cachera MF, Deheeger M, Bellisle F, Sempe M, Guilloud-Bataille M, Patois E. Adiposity rebound in children: a simple indicator for predicting obesity. Am J Clin Nutr 1984;39:129–35.
32. Dietz WH. Critical periods in childhood for the development of obesity. Am J Clin Nutr 1994;59:955–9.
33. Henry CJK, Ulijaszek SJ. Long-term consequences of early environment. Growth, development and the lifespan developmental perspective. Cambridge, UK: Cambridge University Press, 1996.
34. Martorell R, Kettel LK, Schroeder DG. Reversibility of stunting: epidemiological findings in children from developing countries. Eur J Clin Nutr 1994;48(suppl 1): S45–57.
35. Golden MH. Is complete catch-up possible for stunted malnourished children? Eur J Clin Nutr 1994;48(suppl 1):S58–71.
36. Barker DJ. The developmental origins of well-being. Philos Trans R Soc Lond B Biol Sci 2004;359:1359–66.
37. Ong KK, Dunger DB. Perinatal growth failure: the road to obesity, insulin resistance and cardiovascular disease in adults. Best Pract Res Clin Endocrinol Metab 2002;16:191–207.
38. Remacle C, Bieswal F, Reusens B. Programming of obesity and cardiovascular disease. Int J Obes Relat Metab Disord 2004;28(suppl 3):S46–53.

39. Monteiro PO, Victora CG, Barros FC, Monteiro LM. Birth size, early childhood growth, and adolescent obesity in a Brazilian birth cohort. Int J Obes Relat Metab Disord 2003;27:1274–82.
40. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. Bull WHO 1987;65:663–737.
41. Wharton B. Causes of low birthweight in developing countries. In: Senterre J, ed. Intrauterine growth retardation. New York: Raven Press, 1989:143–55.
42. Frongillo EA. Growth of the breast-fed child. In: Martorell R, Haschke F, eds. Nutrition and growth. Philadelphia, Pa, USA: Lippincott Williams and Wilkins, 2001: 37–49.
43. Waterlow JC. Observations on the natural history of stunting. In: Waterlow JC, ed. Linear growth retardation in less developed countries. New York: Raven Press, 1988:1–12.
44. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser 1995;854:1–452.
45. Stoltzfus RJ. Growth of school-age children. In: Martorell R, Haschke F, eds. Nutrition and growth. Philadelphia, Pa, USA: Lippincott Williams and Wilkins, 2001:257–76.
46. Ulijaszek SJ. Ethnic differences in patterns of human growth in stature. In: Martorell R, Haschke F, eds. Nutrition and growth. Philadelphia, Pa, USA: Lippincott Williams and Wilkins, 2001:1–15.
47. Johnston FE, Wainer H, Thissen D, MacVean RB. Hereditary and environmental determinants of growth in height in a longitudinal sample of children and youth of Guatemalan and European ancestry. Am J Phys Anthropol 1976;44:469–76.
48. Hauspie RC, Susanne C. Genetics of child growth. In: Ulijaszek SJ, Johnston FE, Preece MA, eds. Cambridge encyclopaedia of human growth and development. Cambridge, UK: Cambridge University Press, 1998:124–8.
49. Tomkins A. Growth monitoring, screening and surveillance in developing countries. In: Ulijaszek SJ, Mascie-Taylor CGN, eds. Anthropometry: the individual and the population. Cambridge, UK: Cambridge University Press, 1994:108–16.
50. Tomkins AM. Protein-energy malnutrition and risk of infection. Proc Nutr Soc 1986;45:289–304.
51. Steckel RH. Growth depression and recovery: the remarkable case of American slaves. Ann Hum Biol 1987; 14:111–32.
52. Proos LA. Anthropometry in adolescence—secular trends, adoption, ethnic and environmental differences. Horm Res 1993;39(suppl 3):18–24.
53. Sawaya AL, Martins P, Hoffman D, Roberts SB. The link between childhood undernutrition and risk of chronic diseases in adulthood: a case study of Brazil. Nutr Rev 2003;61(5 pt 1):168–75.
54. Ulijaszek SJ, Johnston FE, Preece MA. Cambridge encyclopaedia of human growth and development. Cambridge, UK: Cambridge University Press, 1998.
55. Tobias PV. The negative secular trend. J Hum Evol 1985;14:347–56.
56. Henneberg M, van den Berg ER. Test of socioeconomic causation of secular trend: stature changes among favored and oppressed South Africans are parallel. Am J Phys Anthropol 1990;83:459–65.
57. Ulijaszek SJ. Evidence for a secular trend in heights and weights of adults in Papua New Guinea. Ann Hum Biol 1993;20:349–55.
58. de Castro JJ, Aleixo Dias J, Baptista F, Garcia e Costa J, Galvao-Teles A, Camilo-Alves A. Secular trends of weight, height and obesity in cohorts of young Portuguese males in the District of Lisbon: 1960–1990. Eur J Epidemiol 1998;14:299–303.
59. Rona RJ. Monitoring nutritional status in England and Scotland. In: Hauspie R, Lindgren G, Falkner F, eds. Essays on auxology. Welwyn Garden City, UK: Castlemead, 1995:291–301.
60. Heude B, Lafay L, Borys JM, Thibult N, Lommez A, Romon M, Ducimetiere P, Charles MA. Time trend in height, weight, and obesity prevalence in school children from Northern France, 1992–2000. Diabetes Metab 2003;29:235–40.
61. Rami B, Schober E, Kirchengast S, Waldhor T, Sefranek R. Prevalence of overweight and obesity in male adolescents in Austria between 1985 and 2000. A population based study. J Pediatr Endocrinol Metab 2004;17:67–72.
62. Loesch DZ, Stokes K, Huggins RM. Secular trend in body height and weight of Australian children and adolescents. Am J Phys Anthropol 2000;111:545–56.
63. Lazarus R, Wake M, Hesketh K, Waters E. Change in body mass index in Australian primary school children, 1985–1997. Int J Obes Relat Metab Disord 2000;24:679–84.
64. Thompson AM, Baxter-Jones AD, Mirwald RL, Bailey DA. Secular trend in the development of fatness during childhood and adolescence. Am J Hum Biol 2002;14:669–79.
65. Troiano RP, Flegal KM, Kuczmarski RJ, Campbell SM, Johnson CL. Overweight prevalence and trends for children and adolescents. The National Health and Nutrition Examination Surveys, 1963 to 1991. Arch Pediatr Adolesc Med 1995;149:1085–91.
66. Demerath EW, Li J, Sun SS, Chumlea WC, Remsberg KE, Czerwinski SA, Towne B, Siervogel RM. Fifty-year trends in serial body mass index during adolescence in girls: the Fels Longitudinal Study. Am J Clin Nutr 2004;80:441–6.
67. Reyes MEP, Barahona EEC, Cahuich MB, Barragan A, Malina RM. Growth status of children 6–12 years from two different geographic regions of Mexico. Ann Hum Biol 2002;29:11–25.
68. Hauspie RC, Vercauteren M, Susanne C. Secular changes in growth and maturation: an update. Acta Paediatr Suppl 1997;423:20–7.
69. Bielicki T, Welon Z. Growth data as indicators of social inequalities: the case of Poland. Yearb Phys Anthropol 1982;25:153–67.
70. Blakely T, Hales S, Kieft C, Wilson N, Woodward A. The global distribution of risk factors by poverty level. Bull World Health Organ 2005;83:118–26.
71. Frongillo EA Jr, de Onis M, Hanson KM. Socioeconomic and demographic factors are associated with worldwide patterns of stunting and wasting of children. J Nutr 1997;127:2302–9.
72. Popkin B, Albala C, Benjelloun S, Bourne L, Cannon G, Coitinho D, Doak C, Galal OM, Ghassemi H, Harrison G, Kosulwat V, Lee MJ, Maletnlema T, Matsudo V, Monteiro C, Noor MI, Reddy KS, Rivera J, Rodriguez-Ojea A, Uauy R, Vorster HH, Zhai FY. Part IV. Bellagio Declaration—Nutrition and health transition in the

developing world: the time to act. Public Health Nutr 2002;5:279–80.
73. Smith LC, Haddad L. How important is improving food availability for reducing child malnutrition in developing countries? Agric Econ 2001;26:191–204.
74. Bhattacharya J, Currie J, Haider S. Poverty, food insecurity, and nutritional outcomes in children and adults. J Health Econ 2004;23:839–62.
75. Torun B, Davies PS, Livingstone MB, Paolisso M, Sackett R, Spurr GB. Energy requirements and dietary energy recommendations for children and adolescents 1 to 18 years old. Eur J Clin Nutr 1996;50(suppl 1):S37–81.
76. Golden MH. The role of individual nutrient deficiencies in growth retardation of children as exemplified by zinc and protein. In: Waterlow JC, ed. Linear growth retardation in less developed countries. New York: Raven Press, 1988:143–63.
77. Dewey KG, Beaton G, Fjeld C, Lonnerdal B, Reeds P. Protein requirements of infants and children. Eur J Clin Nutr 1996;50(suppl 1):S119–50.
78. Chen P, Soares AM, Lima AA, Gamble MV, Schorling JB, Conway MV, Barrett LJ, Blaner WS, Guerrant RL. Association of vitamin A and zinc status with altered intestinal permeability: analyses of cohort data from northeastern Brazil. J Health Popul Nutr 2003;21:309–15.
79. Tremblay A. Dietary fat and body weight set point. Nutr Rev 2004;62(7 pt 2):S75–7.
80. Boeing H. Is fat intake related to obesity? What is the evidence from the epidemiological point of view? Ernahrungs Umschau 2005;52:4.
81. Li D, Sinclair AJ, Mann NJ, Turner A, Ball MJ. Selected micronutrient intake and status in men with differing meat intakes, vegetarians and vegans. Asia Pacific J Clin Nutr 2000;9:18–23.
82. Murphy SP, Allen LH. Nutritional importance of animal source foods. J Nutr 2003;133(suppl 2):3932S–5S.
83. Dagnelie PC, van Dusseldorp M, van Staveren WA, Hautvast JG. Effects of macrobiotic diets on linear growth in infants and children until 10 years of age. Eur J Clin Nutr 1994;48(suppl 1):S103–12.
84. Ramakrishnan U, Aburto N, McCabe G, Martorell R. Multimicronutrient interventions but not vitamin A or iron interventions alone improve child growth: results of 3 meta-analyses. J Nutr 2004;134:2592–602.
85. Bhandari N, Bahl R, Taneja S. Effect of micronutrient supplementation on linear growth of children. Br J Nutr 2001;85(suppl 2):S131–7.
86. Rivera JA, Hotz C, Gonzalez-Cossio T, Neufeld L, Garcia-Guerra A. The effect of micronutrient deficiencies on child growth: a review of results from community-based supplementation trials. J Nutr 2003;133(11 suppl 2):4010S–20S.
87. Koletzko B. Response to and range of acceptable fat intakes in infants and children. Eur J Clin Nutr 1999;53(suppl 1):S78–83.
88. Vobecky JS, Vobecky J, Normand L. Risk and benefit of low fat intake in childhood. Ann Nutr Metabol 1995;39:124–33.
89. Nicklas TA, Webber LS, Koschak M, Berenson GS. Nutrient adequacy of low fat intakes for children: the Bogalusa Heart Study. Pediatrics 1992;89:221–8.
90. Tonstad S, Sivertsen M. Relation between dietary fat and energy and micronutrient intakes. Arch Dis Child 1997;76:416–20.
91. Drewnowski A. Obesity and the food environment: dietary energy density and diet costs. Am J Prev Med 2004;27(3 suppl):154–62.
92. Darmon N, Briend A, Drewnowski A. Energy-dense diets are associated with lower diet costs: a community study of French adults. Public Health Nutr 2004;7:21–7.
93. Drewnowski A, Specter SE. Poverty and obesity: the role of energy density and energy costs. Am J Clin Nutr 2004;79:6–16.
94. Dagnelie PC, van Staveren WA, Verschuren SAJM, Hautvast JG. Nutritional status of infants aged 4 to 18 months on macrobiotic diets and matched omnivorous control infants: a population-based mixed-longitudinal study. I. Weaning pattern, energy and nutrient intake. Eur J Clin Nutr 1989;43:311–23.
95. Mueller I, Vounatsou P, Allen BJ, Smith T. Spatial patterns of child growth in Papua New Guinea and their relation to environment, diet, socio-economic status and subsistence activities. Ann Hum Biol 2001;28:263–80.
96. Marreiro DD, Fisberg M, Cozzolino SM. Zinc nutritional status in obese children and adolescents. Biol Trace Elem Res 2002;86:107–22.
97. Liu YX, Li HQ, Yang XQ, Karlberg J. Early linear growth retardation in Chongqing, China. J Paediatr Child Health 1999;35:272–7.
98. Moore SR, Lima AA, Conaway MR, Schorling JB, Soares AM, Guerrant RL. Early childhood diarrhoea and helminthiases associated with long-term linear growth faltering. Int J Epidemiol 2001;30:1457–64.
99. Torres AM, Peterson KE, de Souza AC, Orav EJ, Hughes M, Chen LC. Association of diarrhoea and upper respiratory infections with weight and height gains in Bangladeshi children aged 5 to 11 years. Bull World Health Organ 2000;78:1316–23.
100. Villamor E, Mbise R, Spiegelman D, Hertzmark E, Fataki M, Peterson KE, Ndossi G, Fawzi WW. Vitamin A supplements ameliorate the adverse effect of HIV-1, malaria, and diarrheal infections on child growth. Pediatrics 2002;109:E6.
101. Wilson WM, Dufour DL, Staten LK, Barac-Nieto M, Reina JC, Spurr GB. Gastrointestinal parasitic infection, anthropometrics, nutritional status, and physical work capacity in Colombian boys. Am J Hum Biol 1999;11:763–71.
102. Friedman JF, Kurtis JD, Mtalib R, Opollo M, Lanar DE, Duffy PE. Malaria is related to decreased nutritional status among male adolescents and adults in the setting of intense perennial transmission. J Infect Dis 2003;188:449–57.
103. Bailey RC, Kamenga MC, Nsuami MJ, Nieburg P, St Louis ME. Growth of children according to maternal and child HIV, immunological and disease characteristics: a prospective cohort study in Kinshasa, Democratic Republic of Congo. Int J Epidemiol 1999;28:532–40.
104. de Martino M, Tovo PA, Galli L, Gabiano C, Chiarelli F, Zappa M, Gattinara GC, Bassetti D, Giacomet V, Chiappini E, Duse M, Garetto S, Caselli D; Italian Register for HIV infection in Children. Puberty in perinatal HIV-1 infection: a multicentre longitudinal study of 212 children. AIDS 2001;15:1527–34.
105. Buchacz K, Rogol AD, Lindsey JC, Wilson CM, Hughes MD, Seage GR 3rd, Oleske JM, Rogers AS; Pediatric

AIDS Clinical Trials Group 219 Study Team. Delayed onset of pubertal development in children and adolescents with perinatally acquired HIV infection. J Acquir Immune Defic Syndr 2003;33:56–65.
106. Villamor E, Fataki MR, Bosch RJ, Mbise RL, Fawzi WW. Human immunodeficiency virus infection, diarrheal disease and sociodemographic predictors of child growth. Acta Pediatr 2004;93:372–9.
107. Stephenson L. The impact of schistosomiasis on human nutrition. Parasitology 1993;107(suppl):S107–23.
108. McGarvey ST, Wu G, Zhang S, Wang Y, Peters P, Olds GR, Wiest PM. Child growth, nutritional status and Schistosomiasis japonica in Jiangxi, People's Republic of China. Am J Trop Med Hyg 1993;48:547–53.
109. Checkley W, Gilman RH, Epstein LD, Suarez M, Diaz JF, Cabrera L, Black RE, Sterling CR. Asymptomatic and symptomatic cryptosporidiosis: their acute effect on weight gain in Peruvian children. Am J Epidemiol 1997; 145:156–63.
110. Checkley W, Epstein LD, Gilman RH, Black RE, Cabrera L, Sterling CR. Effects of Cryptosporidium parvum infection in Peruvian children: growth faltering and subsequent catch-up growth. Am J Epidemiol 1998;148:497–506.
111. Lacey BE, Rosemore J. Helicobacter pylori: ulcers and more: the beginning of an era. J Nutr 2001;131:2789S–93S.
112. Jolobe OMP. Helicobacter pylori infection with iron deficiency anaemia and subnormal growth at puberty. Arch Dis Child 2000;82:428.
113. Buyukgebiz A, Dundar B, Bober E, Buyukgebiz B. Helicobacter pylori infection in children with constitutional delay of growth and puberty. J Pediatr Endocrinol Metab 2001;14:549–51.
114. Tomkins A. Nutrition, infection and immunity: public health implications. In: Calder PC, Field CJ, Gill HS, eds. Nutrition and immune function. Wallingford, Oxfordshire, UK: CABI Publishing, 2002.
115. Simon D. Puberty in chronically diseased patients. Horm Res 2002;57(suppl 2):53–6.
116. Farmer P. Sidney W. Mintz lecture for 2001. An anthropology of structural violence. Curr Anthropol 2004;45:305–25.
117. Walls T, Shingadia D. Global epidemiology of paediatric tuberculosis. J Infect 2004;48:13–22.
118. Bates I, Fenton C, Gruber J, Lalloo D, Medina Lara A, Squire SB, Theobald S, Thomson R, Tolhurst R. Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part I: Determinants operating at individual and household level. Lancet Infect Dis 2004;4:267–77.
119. Malcolm CA, MacKay WG, Shepherd A, Weaver LT. Helicobacter pylori in children is strongly associated with poverty. Scott Med J 2004;49:136–8.
120. Semba RD, Miotti P, Chiphangwi JD, Henderson R, Dallabetta G, Yang LP, Hoover D. Maternal vitamin A deficiency and child growth failure during human immunodeficiency virus infection. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:219–22.
121. Schell LM. Effects of pollutants on human prenatal and postnatal growth: noise, lead, polychlorobiphenyl compounds, and toxic wastes. Yearb Phys Anthropol 1991;34:157–88.
122. Paigen B, Goldman LR, Magnant MM, Highland JH, Steegman AT Jr. Growth of children living near the hazardous waste site, Love Canal. Hum Biol 1987;59: 489–508.
123. Schell L, Cameron N. Weight growth velocity from birth to 2 years of age in relation to lead burden. Am J Phys Anthropol 2003;(suppl 36):185.
124. Masuda Y, Yoshimura H. Polychlorinated biphenyls and dibenzofurans in patients with yusho and their toxicological significance: a review. Am J Ind Med 1984;5:31–44.
125. Yoshimura T, Ikeda M. Growth of school children with polychlorinated biphenyl poisoning or yusho. Environ Res 1978;17:416–25.
126. Huh CA, Chen HY. History of lead pollution recorded in East China Sea sediments. Mar Pollut Bull 1999;38: 545–9.
127. Chu NF, Liou SH, Wu TN, Ko KN, Chang PY. Risk factors for high blood lead levels among the general population in Taiwan. Eur J Epidemiol 1998;14:775–81.
128. Park DL. Effect of processing on aflatoxin. Adv Exp Med Biol 2002;504:173–9.
129. Hell K, Cardwell KF, Poehling HM. Relationship between management practices, fungal infection and aflatoxin for stored maize in Benin. J Phytopathol 2003;151:690–8.
130. Bankole SA, Mabekoje OO. Mycoflora and occurrence of aflatoxin B1 in dried yam chips from markets in Ogun and Oyo States, Nigeria. Mycopathologia 2004;157:111–5.
131. Bankole SA, Ogunsanwo BM, Eseigbe DA. Aflatoxins in Nigerian dry-roasted groundnuts. Food Chem 2005; 89:503–6.
132. Kamkar A. A study on the occurrence of aflatoxin M1 in raw milk produced in Sarab city of Iran. Food Control 2005;16:593–9.
133. Rastogi S, Dwivedi PD, Khanna SK, Das M. Detection of aflatoxin M1 contamination in milk and infant milk products from Indian markets by ELISA. Food Control 2004;15:287–90.
134. Yamashita A, Yoshizawa T, Aiura Y, Sanchez PC, Dizon EI, Arim RH. Fusarium mycotoxins (fumonisins, nivalenol, and zearalenone) and aflatoxins in corn from Southeast Asia. Biosci Biotechnol Biochem 1995;59:1804–7.
135. Scussel VM. Aflatoxin and food safety: recent South American perspectives. J Toxicol Toxin Rev 2004;23: 179–216.
136. Bittencourt ABF, Oliveira CAF, Dilkin P, Correa B. Mycotoxin occurrence in corn meal and flour traded in Sao Paulo, Brazil. Food Control 2005;16:117–20.
137. Sales AC, Yoshizawa T. Mold counts and Aspergillus section Flavi populations in rice and its by-products from the Philippines. J Food Prot 2005;68:120–5.
138. Lin LC, Liu FM, Fu YM, Shih DYC. Survey of aflatoxin M-1 contamination of dairy products in Taiwan. J Food Drug Anal 2004;12:154–60.
139. European Commission. Key Action 1 (KA1) on Food, Nutrition and Health, QLK1-CT-2000-01248. Brussels, Belgium: Quality of Life and Management of Living Resources Programme, 2000.
140. European Commission. Risk assessment of aflatoxins. Reports on tasks for scientific cooperation. Report EUR 17526. Brussels: European Commission, 1997.
141. Wood GE. Mycotoxins in foods and feeds in the United States. J Anim Sci 1992;70:3941–9.

142. Mendez-Albores JA, Arambula-Villa G, Loarca-Pina MG, Gonzalez-Hernandez J, Castano-Tostado E, Moreno-Martinez E. Aflatoxins' fate during the nixtamalization of contaminated maize by two tortilla-making processes. J Stored Prod Res 2004;40:87–94.
143. Plasencia J. Aflatoxins in maize: a Mexican perspective. J Toxicol Toxin Rev 2004; 23:155–77.
144. Nordstrom-Klee B, Delaney-Black V, Covington C, Ager J, Sokol R. Growth from birth onwards of children prenatally exposed to drugs: a literature review. Neurotoxicol Teratol 2002;24:481–8.
145. Lichty JA, Ting RY, Bruns PD, Dyar E. Studies of babies born at high altitudes. I. Relation of altitude to birth weight. AMA J Dis Child 1957;93:666–9.
146. Haas JD, Frongillo EA, Stepick CD, Beard JL, Hurtado G. Altitude, ethnic and sex difference in birth weight and length in Bolivia. Hum Biol 1980;52:459–77.
147. Frisancho AR. Human adaptation and accommodation. Ann Arbor, Mich, USA: University of Michigan Press, 1993.
148. Demeer K, Bergman R, Kusner JS, Voorhoeve HWA. Differences in physical growth of Aymara and Quechua children living at high-altitude in Peru. Am J Phys Anthropol 1993;90:59–75.
149. Weitz CA, Garruto RM. Growth of Han migrants at high altitude in Central Asia. Am J Hum Biol 2004;16:405–19.
150. Stinson S. The effect of high altitude on the growth of children of high socioeconomic status in Bolivia. Am J Phys Anthropol 1982;59:61–71.
151. Greksa LP, Spielvogel H, Paredes-Fernandez L, Paz-Zamora M, Caceres E. The physical growth of urban children at high altitude. Am J Phys Anthropol 1984;65:315–22.
152. Barker DJ, Forsen T, Uutela A, Osmond C, Eriksson JG. Size at birth and resilience to effects of poor living conditions in adult life: longitudinal study. BMJ 2001;323:1273–6.
153. Williams S, Dickson N. Early growth, menarche, and adiposity rebound. Lancet 2002;359(9306):580–1.
154. Janz KF, Levy SM, Burns TL, Torner JC, Willing MC, Warren JJ. Fatness, physical activity, and television viewing in children during the adiposity rebound period: the Iowa Bone Development Study. Prev Med 2002;35:563–71.
155. Cameron N, Demerath EW. Critical periods in human growth and their relationship to diseases of aging. Am J Phys Anthropol 2002;(suppl 35):159–84.
156. Boudailliez B, Fremaux MP, Jeanne F, Escoffier I, Bony H. Adolescent obesity: guidelines for the management. Arch Pediatr 2004;11:1274–6 (in French).
157. Vazir S. Behavioral aspects of development of eating behaviour and nutrition status. Nutr Rev 2002;60(5 pt 2):S95–101.
158. Guldan GS, Zhang MY, Zhang YP, Hong JR, Zhang HX, Fu SY, Fu NS. Weaning practices and growth in rural Sichuan infants: a positive deviance study. J Trop Pediatr 1993;39:168–75.
159. Powell CA, Walker SP, Chang SM, Grantham-McGregor SM. Nutrition and education: a randomized trial of the effects of breakfast in rural primary school children. Am J Clin Nutr 1998;68:873–9.
160. Little MA, Gray SJ, Leslie PW. Growth of nomadic and settled Turkana infants of northwest Kenya. Am J Phys Anthropol 1993;92:273–89.
161. Gray SJ, Wiebusch B, Akol HA. Cross-sectional growth of pastoralist Karimojong and Turkana children. Am J Phys Anthropol 2004;125:193–202.
162. Little MA. Adaptability of African pastoralists. In: Ulijaszek SJ, Huss-Ashmore RA, eds. Human adaptability. Past, present and future. Oxford, UK: Oxford University Press, 1997:29–60.
163. Abbie A. Studies in physical anthropology: the Aboriginal Growth Pattern. Canberra: Australian Institute of Aboriginal Studies, 1975.
164. Henneberg M, Schilitz A, Lambert KM. Assessment of the growth of children and physical status of adults in two Aboriginal communities in South Australia. Am J Hum Biol 2001;13:603–11.
165. Daniel M, Rowley KG, McDermott R, O'Dea K. Diabetes and impaired glucose tolerance in Aboriginal Australians: prevalence and risk. Diabetes Res Clin Pract 2002;57:23–33.
166. Norgan NG. Population differences in body composition in relation to the body mass index. Eur J Clin Nutr 1994;48(suppl 3):S10–27.
167. Phillips NT, Henneberg M, Norgan N, Schmitt L, Ulijaszek SJ. Obesity, growth and body proportionality in Australian Aboriginal children. Australasian Society for Human Biology Eighteenth Annual Conference, Canberra, 2004.
168. Bailey RC. The comparative growth of Efe pygmies and African farmers from birth to age 5 years. Ann Hum Biol 1991;18:113–20.
169. Merimee TJ, Zapf J, Hewlett B, Cavalli-Sforza LL. Insulin-like growth factors in pygmies. The role of puberty in determining final stature. N Engl J Med 1987;316:906–11.
170. Geffner ME, Bersch N, Bailey RC, Golde DW. Growth hormone induces resistance to the mitogenic action of insulin through local IGF-I. Studies in normal and pygmy T-cell lines. Diabetes 1994;43:68–72.
171. Hattori Y, Vera JC, Rivas CI, Bersch N, Bailey RC, Geffner ME, Golde DW. Decreased insulin-like growth factor I receptor expression and function in immortalized African pygmy T cells. J Clin Endocrinol Metabol 1996;81:2257–63.
172. Widdowson EM. Mental contentment and physical growth. Lancet 1951;1(24):1316–8.
173. Montgomery SM, Bartley MJ, Wilkinson RG. Family conflict and slow growth. Arch Dis Child 1997;77: 326–30.
174. Albanese A, Hamill G, Jones J, Skuse D, Matthews DR, Stanhope R. Reversibility of physiological growth hormone secretion in children with psychosocial dwarfism. Clin Endocrinol 1994;40:687–92.
175. Skuse D, Albanese A, Stanhope R, Gilmour J, Voss L. A new stress-related syndrome of growth failure and hyperphagia in children, associated with reversibility of growth-hormone insufficiency. Lancet 1996;348(9024):353–8.
176. Gilmour J, Skuse D. A case-comparison study of the characteristics of children with a short stature syndrome induced by stress (hyperphagic short stature) and a consecutive series of unaffected "stressed" children. J Child Psychol Psychiatr 1999;40:969–78.
177. Gohlke BC, Khadilkar VV, Skuse D, Stanhope R. Recognition of children with psychosocial short stature:

a spectrum of presentation. J Pediatr Endocrinol Metab 1998;11:509–17.
178. Skuse DH. Growth and psychosocial stress. In: Ulijaszek SJ, Johnston FE, Preece MA, eds. Cambridge encyclopedia of human growth and development. Cambridge, UK: Cambridge University Press, 1998:341–2.
179. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. Bull World Health Organ 1987;65:663–737.
180. Cole TJ. Seasonal effects on physical growth and development. In: Ulijaszek SJ, Strickland SS, eds. Seasonality and human ecology. Cambridge, UK: Cambridge University Press, 1993:89–106.
181. Schnaas L, Rothenberg SJ, Flores MF, Martinez S, Hernandez C, Osorio E, Perroni E. Blood lead secular trend in a cohort of children in Mexico City (1987–2002). Environ Health Perspect 2004;112:1110–5.
182. Eveleth PB, Tanner JM. Worldwide variation in human growth, 2nd ed. Cambridge, UK: Cambridge University Press, 1990.
183. Berkey CS, Gardner JD, Frazier AL, Colditz GA. Relation of childhood diet and body size to menarche and adolescent growth in girls. Am J Epidemiol 2000;152:446–52.
184. Eneobong HN, Akosa UM. Adolescents living in boarding houses in Nsukka, Enugu State, Nigeria. 1. Meal patterns, nutrition knowledge and nutrient intake. Ecol Food Nutr 1993;30:179–93.
185. Ulijaszek SJ. Immunocompetence. In: Ulijaszek SJ, Johnston FE, Preece MA, eds. Cambridge encyclopaedia of human growth and development. Cambridge, UK: Cambridge University Press, 1998:340.
186. Choe YH, Kim SK, Hong YC. Helicobacter pylori infection with iron deficiency anaemia and subnormal growth at puberty. Arch Dis Child 2000;82:136–40.
187. Blanck HM, Marcus M, Tolbert PE, Rubin C, Henderson AK, Hertzberg VS, Zhang RH, Cameron L. Age at menarche and Tanner stage in girls exposed in utero and postnatally to polybrominated biphenyl. Epidemiology 2000;11:641–7.
188. Den Hond E, Roels HA, Hoppenbrouwers K, Nawrot T, Thijs L, Vandermeulen C, Winneke G, Vanderschueren D, Staessen JA. Sexual maturation in relation to polychlorinated aromatic hydrocarbons: Sharpe and Skakkebaek's hypothesis revisited. Environ Health Perspect 2002;110:771–6.
189. Denham M, Schell LM, Deane G, Gallo MV, Ravenscroft J, DeCaprio AP; Akwesasne Task Force on the Environment. Relationship of lead, mercury, mirex, dichlorodiphenyldichloroethylene, hexachlorobenzene, and polychlorinated biphenyls to timing of menarche among Akwesasne Mohawk girls. Pediatrics 2005;115: e127–34.
190. Tren R, Bate R. Malaria and the DDT story. London: Institute of Economic Affairs, 2001.
191. Gladen BC, Klebanoff MA, Hediger ML, Katz SH, Barr DB, Davis MD, Longnecker MP. Prenatal DDT exposure in relation to anthropometric and pubertal measures in adolescent males. Environ Health Perspect 2004;112:1761–7.
192. Patton GC, McMorris BJ, Toumbourou JW, Hemphill SA, Donath S, Catalano RF. Puberty and the onset of substance use and abuse. Pediatrics 2004;114:e300–6.
193. US Department of Health and Human Services. The health consequences of smoking for women: a report of the Surgeon General. Rockville, Md, USA: US Department of Health and Human Services 1980;217–21.
194. Luciano A, Bolognani M, Biondani P, Ghizzi C, Zoppi G, Signori E. The influence of maternal passive and light active smoking on intrauterine growth and body composition of the newborn. Eur J Clin Nutr 1998;52:760–3
195. Charlton A. Children and smoking: the family circle. Br Med Bull 1996;52:90–107.
196. Golub MS. Adolescent health and the environment. Environ Health Perspect 2000;108:355–62.
197. Greksa LP. Age of menarche in Bolivian girls of European and Aymara ancestry. Ann Hum Biol 1990;17:49–53.
198. Frisancho AR, Baker PT. Altitude and growth: a study of the patterns of physical growth of a high-altitude Peruvian Quechua population. Am J Phys Anthropol 1970;32:279–92.
199. Clegg EJ, Pawson IG, Ashton EH, Flinn RM. The growth of children at different altitudes in Ethiopia. Philos Trans R Soc Lond B Biol Sci 1972;264:403–37.
200. McCammon RW. Human growth and development. Springfield, Ill, USA: Charles C Thomas, 1970.
201. Belsky J, Steinberg L, Draper P. Childhood experience, interpersonal development, and reproductive strategy: an evolutionary theory of socialization. Child Dev 1991;62:647–70.
202. Kim K, Smith PK, Palermiti AL. Conflict in childhood and reproductive development. Evol Hum Behav 1997;18:109–42.
203. Hulanicka B. Acceleration of menarcheal age of girls from dysfunctional families. J Reprod Infant Psychol 1999;17:119–32.
204. Hulanicka B, Gronkiewicz L, Koniarek J. Effect of familial distress on growth and maturation of girls: a longitudinal study. Am J Hum Biol 2001;13:771–6.

Physical activity and fitness in an international growth standard for preadolescent and adolescent children

Robert M. Malina and Peter T. Katzmarzyk

Abstract

Concepts related to energy expenditure, physical activity and physical fitness, and methods of assessment are briefly considered. Variation in energy expenditure, physical activity, and physical fitness associated with age and sex during childhood and adolescence and relationships between physical activity and physical fitness in children and adolescents are reviewed. Implications of undernutrition and obesity for physical activity and physical fitness, and secular changes in physical activity and physical fitness, are briefly highlighted. The review concludes with specific recommendations for and limitations of inclusion of indicators of physical activity and fitness in the construction of an International Growth Standard for Preadolescent and Adolescent Children.

Introduction

The prevalence of obesity has increased among youths worldwide. Reduced levels of habitual physical activity are hypothesized to be a factor in the increased prevalence of obesity [1]. Obesity, physical inactivity, and poor physical fitness are independent risk factors for chronic disease as well as premature mortality among adults [2–4]. The increased prevalence of these risk factors among children and adolescents and the emergence of symptoms of metabolic and cardiovascular diseases during childhood and adolescence highlight the need to consider physical activity and physical fitness in the development of an International Growth Standard for Preadolescent and Adolescent Children.

Robert M. Malina is Research Professor at Tarleton State University, Stephenville, Texas, USA; Peter T. Katzmarzyk is affiliated with the School of Physical and Health Education, Department of Community Health and Epidemiology, Queen's University, Kingston, Ontario, Canada.

Please direct queries to the corresponding author: Robert M. Malina, 10735 FM 2668, Bay City, TX 77414, USA; e-mail: rmalina@wcnet.net.

Physical activity

Physical activity is a behavior involving movement of the body through space. It has several dimensions. Physical activity is viewed most often in terms of energy expenditure and the stresses and strains associated with weight-bearing and ground-reaction forces. It also has a major performance component viewed primarily in specific movement skills and measures of physical fitness. Context refers to the settings and types of physical activities (sport, play, education, work, "exercise," etc.) and is strongly influenced by culture.

Physical fitness

Physical fitness is a state or a condition that permits the individual to carry out his or her daily activities without undue fatigue and with sufficient reserve to enjoy active leisure pursuits. It is often an assumed correlate of physical activity though relationships are generally moderate (see below). Physical fitness was historically viewed in terms of three components: muscular strength and endurance, cardiorespiratory endurance, and motor ability [5]. Over the past 50 to 60 years, surveys of the physical fitness of youths have focused largely on performances in a variety of strength and motor tasks (performance-related fitness) and smaller-scale studies of cardiorespiratory endurance. The concept of physical fitness has since evolved from a primary focus on motor and strength components to more emphasis on health (health-related physical fitness) in the late 1970s [6]. Although specific tests vary, health-related fitness includes tests of cardiorespiratory endurance, muscular strength and endurance, musculoskeletal function of the lower trunk and upper thighs, and/or body composition, specifically fatness [6–8]. The concept of physical fitness continues to evolve. Morphological and metabolic components have been added to the more traditional muscular strength and endurance, motor, and cardiovascular components [9].

Measurement of physical activity

A summary of methods commonly used to estimate habitual levels of physical activity and energy expenditure, including advantages and limitations, in children and adolescents is given in **table 1**. Detailed consideration of each method is beyond the scope of this review. None of the methods covers all aspects of physical activity and energy expenditure; in turn, a combination of methods is needed to obtain a comprehensive estimate of habitual level of physical activity and energy expenditure of an individual [10, 11]. The choice of combinations depends on the specific objectives of a study, age of subjects, and availability of equipment and personnel. Questionnaires are the instrument of choice for large-scale surveys of physical activity, but the scope and focus of questionnaires vary among studies. Most instruments also include indicators of physical inactivity, e.g., television viewing and video games, among others. Large-scale surveys date to the 1980s; several examples are summarized in **table 2**. Several earlier European studies focused on adolescents [23–26].

Measurement of physical fitness

The assessment of physical fitness has a longer history than that of physical activity. Components of physical fitness have traditionally included strength, endurance, speed, power, agility, flexibility, and coordination; morphological and metabolic components have been added more recently. Several physical fitness batteries are summarized in **table 2**. Nationally based standardized fitness test batteries, based primarily on performance-

TABLE 1. Commonly used methods for the assessment of pattern and/or level of physical activity (PA) and energy expenditure (EE)

Method	Function assessed	Advantages	Drawbacks	Comments
Questionnaire	PA	Simple, low cost; suitable for large-scale studies	Relies on memory; hard to quantify; low validity	The shorter the recall period, the higher the validity
Interview	PA	More valid than a questionnaire	Relies on memory	Interviewer can corroborate information
Diary	PA	Short recall time	Interactive	Depends on child's interpretation
Direct observation	PA, (EE?)	No need for recall; context documented	Expensive; depends on observer's skill	"Gold standard" for specific behavioral aspects of activity
Time-lapse (or video) photography	PA, (EE?)	Objective, hard record available	Child is limited to predetermined area	Less expensive than direct observation
Movement counters	PA, (EE?)	Objective, little interaction; low cost	Do not detect specific movements	
Accelerometry	PA, (EE?)	Same as counters, plus acceleration	Does not detect specific activities	Some validity vs. measurements of EE
Heart-rate monitoring	EE	Little interaction; inexpensive	Heart rate affected not only by metabolism	Needs individual calibration vs. VO_2
VO_2—metabolic cart	EE	Measures metabolism	Limited activities; need for mouthpiece or face mask	Useful for ergometry and VO_2-heart rate calibration
VO_2—portable equipment	EE	Measures metabolism away from the laboratory	Highly interactive; expensive	Limited pediatric use in prolonged observations
VO_2—canopy	EE	Measures metabolism	RMR only	Used in conjunction with heart-rate monitoring
Respiration chamber	EE	Precise measurement of EE	Very limited quarters; expensive	Validating other tests; ideal for BMR
Doubly labeled water	EE	Best measure of EE; not interactive	Very high cost; requires at least 1 week	"Gold standard" for average EE, but not for profile of EE

RMR, resting metabolic rate; BMR, basal metabolic rate; VO_2, oxygen uptake; ? = denotes uncertain validity.
Source: [10]. Reproduced with permission from RM Malina, C Bouchard, O Bar-Or. Growth, maturation, and physical activity. 2nd ed. Champaign, IL: Human Kinetics, 2004.

related items, were administered to representative samples of school-aged children in the United States in the 1950s, 1960s, 1970s, and 1980s [27, 28, 35–37] and also to Canadian schoolchildren in the 1960s [31]. Health-related fitness tests were used in surveys of US youths 10 to 17 years of age [7] and in youths grades 1 through 12 [12, 13]. Health-related fitness tests were also used in national surveys of the Canadian population 7 years of age and older, the Canada Fitness survey in 1981 [8] and Campbell's Survey of Well-Being in 1988 [15]. National data for PWC_{170} (physical working capacity at a heart rate of 170 beats per minute), now an indicator of health-related fitness, were reported for Canadian school-aged children in the 1960s [32] and 1980s [38].

A variety of test batteries were used in Europe, with that used in the Leuven Longitudinal Study of Belgian Boys perhaps the most systematically developed [39–41]. The extensive Belgian studies were used in part as the basis for the EUROFIT test battery [33]. Combinations of national and international test bat-

TABLE 2. Selected surveys of physical activity.

Survey	Age	Items
National Children and Youth Fitness Study I, 1984 [12]	Grades 5–12	Activities in school physical education Physical activity outside of school physical education
National Children and Youth Fitness Study II, 1986 [13]	Grades 1–4	Activities in school physical education Activities at home and community (parental report) Parental activity
CDC Youth Risk Behavior Surveillance System [14]	Grades 9–12	Participation in physical activity Enrollment in physical education, activity in physical education Participation in sports (in and out of school) Television viewing
Canada Fitness Survey, 1981 [8], Campbell's Survey on Well-Being, 1988 [15]	10+ yr	Participation in leisure-time physical activities Specific activities coded on a detailed questionnaire according to intensity, duration, and frequency
WHO Cross-National Survey, 1983-1984; WHO Collaborative Survey, 1985-1986 [16]	11, 13, 15 yr	Items dealing with exercise and leisure-time activities: Exercise—times and hours per week Participation in sports clubs Reasons for participating in activity Leisure activities—television, movies or videos, computer games
Chino-Japanese Cooperative Study [17, 18]	7-20 yr	Time spent exercising and/or practicing sports Time spent walking or cycling to school Time spent in additional classes Time spent on music, art, etc. Time spent viewing television
New South Wales Schools Fitness and Activity Survey [19]	2nd, 4th, 6th, 8th, 10th years in school	Participation in organized sports, games, and other activities Participation in nonorganized activities Physical education—periods per week, time in vigorous activity Sedentary recreation—television, videos, computer games
WHO Health Behavior in School-Aged Children Survey, 1985/86, 1989/90, 1993/94, 1997/98, 2001/02 [20]	11, 13, 15 yr	Items dealing with exercise and leisure-time activities: Exercise—times and hours per week Sedentary leisure activities—television, video and computer use, homework
Canadian National Population Health Survey, 1994, 1996, 1998; Canadian Community Health Survey, 2000, 2003 [21]	12+ yr	Participation in leisure-time physical activities Specific activities coded on a detailed questionnaire according to intensity, duration, and frequency
CDC Global School-Based Student Health Survey [22]	13–15 yr	Days physically active at least 60 min per day during the past 7 days and during a typical week Time spend sitting and watching television, playing computer games, talking with friends, etc., during a typical or usual day Days walking or riding a bicycle to school during past 7 days Time usually taken to get to and from school each day

teries have been used in a number of regional and/or national physical fitness surveys, e.g., Australia [19, 42], Hong Kong [34], Belgium [43], Poland [44], Venezuela [45], Guatemala [46], Colombia [47], Argentina [48], China and Japan [17, 18], and Taiwan [162].

Morphological fitness includes indicators of fatness (body-mass index (BMI), subcutaneous and visceral fat, abdominal circumference, and relative fat distribution) as well as bone health and joint flexibility. Most surveys of school-aged youths include measures of height and weight (expressed primarily as the BMI) and occasionally skinfold thickness (see **table 3**). Metabolic fitness includes measures of serum lipids (cholesterol and its fractions), triglycerides, blood pressures, blood glucose, glycemia, and other risk factors for disease. Few national surveys of physical fitness include indicators of metabolic fitness. The Australian Health and Fitness Survey 1985 is an exception; it included fasting total and high-density lipoprotein (HDL) cholesterol and triglycerides in youths 9, 12, and 15 years of age [42]. Morphological and metabolic fitness will not be considered in this report. Obesity, or excessive adiposity, is an important aspect of morphological fitness.

Energy expenditure and physical activity in childhood and adolescence

Estimated 24-hour energy expenditure (EE, kcal/kg) based on doubly labeled water declines with age, beginning as early as 4 to 5 years of age [49]. The decline is especially apparent during the second decade of life. Estimated EE is, on average, greater in males than in females, and the sex difference increases with age. Estimates for longitudinal samples of Dutch children

TABLE 3. Selected tests of physical fitness

Test	Age	Test items
AAHPER Youth Fitness Test [27, 28]	9–17+ yr	Speed—50-yard dash Power/coordination—standing long jump Speed/agility—shuttle run Upper body functional strength—pullups (boys), flexed arm hang (girls) Abdominal strength/endurance—situps Power/coordination—distance throw (softball) Cardiovascular endurance—600-yard run
AAHPERD Health-Related Physical Fitness Test [6, 7]	5–17+ yr	Cardiovascular endurance—1-mile run, 9-min run; 1.5-mile run, 12-min run (13–18 yr) Abdominal strength/endurance—situps Flexibility, lower trunk—sit and reach Subcutaneous fatness—sum triceps + subscapular skinfolds
President's Council on Physical Fitness and Sports [37]	6–17 yr	Upper arm-shoulder girdle strength and endurance—pullups, flexed arm hang Abdominal strength and endurance—curl- ups Lower limb muscle strength, endurance and agility—shuttle run Explosive power of lower limbs—standing long jump Running speed—50-yard dash Cardiorespiratory endurance—1-mile run/walk, 2-mile walk Hamstring, low back flexibility—V sit and reach
Prudential Fitnessgram [29, 30]	5–17+	Body composition—% fat (from triceps + subscapular skinfolds), BMI Aerobic capacity—1-mile run, multistage 20-m shuttle run (PACER) Abdominal strength—curl-up Upper body strength—pushup, modified pullup, pullup, flexed arm hand Trunk extensor strength/flexibility— trunk lift Flexibility—sit and reach, shoulder stretch
CAHPER Fitness Performance Test [31, 32]	7–17 yr	Speed—50-yard dash Speed/agility—shuttle run Power/coordination—standing long jump Abdominal strength/endurance—speed situps Upper body strength—flexed arm hang Endurance—300-yard run Aerobic capacity—PWC_{170}
Canada Fitness Survey, 1981 [8], Campbell's Survey of Well-Being, 1988 [15]	7+ yr	Aerobic fitness—step test Strength—right and left grip Upper body functional strength—push-ups Abdominal strength/endurance—speed situps Trunk flexibility—sit and reach

continued

and adolescents 6 to 17 years of age based on heart rate monitoring indicate similar trends [50, 51].

Energy expenditure associated with physical activity is the most variable component of EE. The ratio of total EE (TEE) to resting EE (REE) is used to estimate the contribution of activity-related energy expenditure to TEE over 24 hours. It is expressed as the "physical activity level" (PAL). The PAL increases during the first 2 years of life [52]. PAL increases with age during childhood and adolescence in children from industrialized countries, but more so in those from rural areas and cities of developing countries [49]. Active children have a PAL of about 1.7 to 2.0.

Estimates of physical activity based on a variety of techniques indicate rather stable levels or a slight increase during childhood and a subsequent decline during the second decade. Activities of children 3 to 10 years of age tend to be spontaneous, largely non-organized, and composed of intermittent brief bouts. Direct observations (minute by minute) of 6- to 10-year-old boys and girls indicate the primacy of intermittent activities of short duration [53]. The median duration for light-to-moderate activity bouts was 6 seconds; most high-intensity bouts did not exceed 3 seconds, and 95% lasted less than 15 seconds. In contrast, activities of older children and adolescents tend to be more organized and of a more regular, prolonged nature.

Most physical activity data are available for children and adolescents 10 years of age and older and are based largely on questionnaires and interviews. Activity level appears to peak at about 12 to 14 years and then, on average, declines across adolescence. The timing and magnitude of the decline varies among studies and according to the context of physical activity. The results from two studies highlight general trends. The US 1992 Youth Risk Behavior Survey supplement suggests the following trends between 12 and 21 years [54]: regular vigorous activity (≥ 3 days/week of running, jogging, or swimming) declined by 29% in males and 36% in females; regular sustained, light-to-moderate activity (≥ 5 days/week and 30+ minutes/occasion of walking or bicycling) declined by 16% in males and 10% in females; strengthening activities (≥ 3 days/week of strengthening or toning exercises) declined by 19% in males and 21% in females; and stretching activities (≥ 3 days/week of stretching exercises) declined by 17% in males and 28% in females. Overall, males were more active than females, but the magnitude of the difference between the sexes varied with activity: the difference was 11.3% for regular vigorous activity, 5.4% for regular

TABLE 3. Selected tests of physical fitness *(continued)*

Test	Age	Test items
European Test of Physical Fitness (EUROFIT) [33]	6–18 yr	Cardiorespiratory endurance—endurance shuttle run, PWC_{170} Balance—flamingo stand Speed of limb movement—plate tapping Flexibility—sit and reach Explosive strength—standing long jump Static strength—hand grip Abdominal strength/endurance—situps Arm/shoulder endurance—flexed arm hang Speed/agility—shuttle run
Chino-Japanese Cooperative Study [17, 18]	7–18 yr	Strength–back, grip Flexibility—standing trunk flexion Abdominal strength/endurance—situps Power—vertical jump Speed/agility—shuttle run Upper body functional strength—pull-ups Power/coordination—hand ball throw Speed—50-m dash Endurance—5-min run
Asia Council for the Standardization of Physical Fitness Tests [34]	9–18 yr	Cardiorespiratory endurance—distance runs: 1,000 m (boys), 800 m (girls), 600 m (< 11 yr) Speed/power—50-m dash Agility/power—shuttle run Leg power—standing long jump Muscular strength—pullups (boys), flexed arm hang (girls, children < 11 yr), flexed leg situps, grip strength Trunk flexibility—forward trunk flexion
Taiwan 1997 Nationwide Children and Youth Fitness Study [162]	7-18 yr	Abdominal strength and endurance—situps Leg power—standing long jump Flexibility—sit and reach Cardiorespiratory endurance—800/1,600-meter walk/run

BMI, body-mass index; PWC_{170}, physical working capacity at a heart rate of 170 beats per minute

sustained activity, 18.2% for strengthening activity, and none for stretching.

The Amsterdam Longitudinal Growth and Health Study followed adolescents from 13 to 16 years of age and then again at 21 years [55]. Overall, males were more active than females, but the difference between the sexes in total activity (> 4 MET [metabolic equivalent] min/week) declined with age, was negligible at 16 years, and showed little change to 21 years. There was variation in age- and sex-associated variation by context: activity in organized sports (sport clubs) showed no sex or age difference over adolescence (approximately 80 to 90 min/week); non-organized sport activity (leisure time sport) showed a small difference between the sexes and declined with age from 13 to 16 years in males (approximately 260 to 140 min/week) but fluctuated in females (approximately 200 to 150 min/week), and declined at 21 years (< 40 min/week in both sexes); and all other activity (school, work, active transport, non-sport leisure) showed no difference between the sexes, was stable at 13 to 16 years (approximately 300 min/week), and increased at 21 years (approximately 425 min/week).

Recent estimates of the percentages of 13- to 15-year-old youths from several countries who were physically active for 7 days per week and who spent 3 or more hours in sedentary activities on a typical day are shown in **table 4**. These are early data from the Global School-Based Student Health Survey [22]. The percentages of 13-year-old boys and girls in the United States, Canada, and Europe meeting recommended levels of physical activity, defined as at least 1 hour of moderate-intensity activity on 5 or more days per week, are shown in **figures 1** and **2**, respectively. The data are

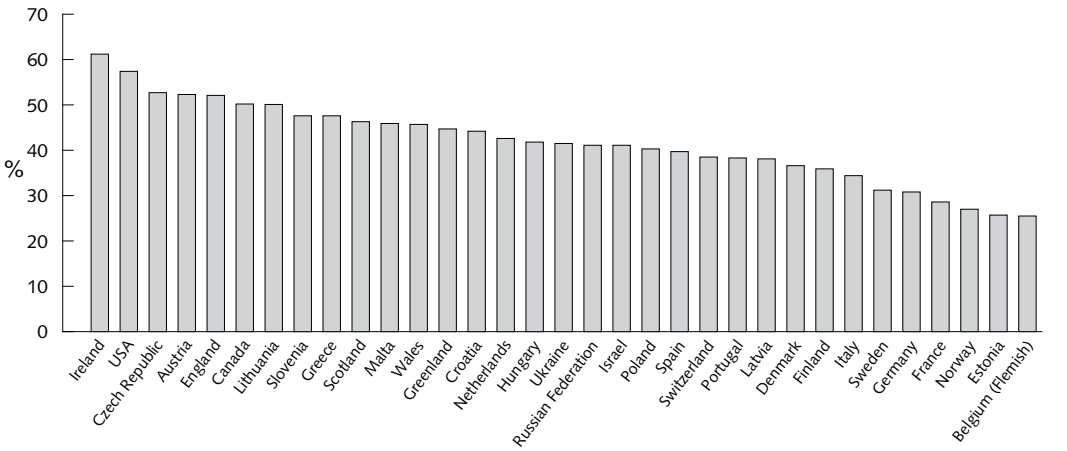

FIG. 1. Proportion of 13-year-old boys meeting recommended levels of physical activity, defined as at least 1 hour of moderate-intensity activity on 5 or more days of the week. Data from the 2001/2002 Health Behaviour in School Aged Children Survey [20]

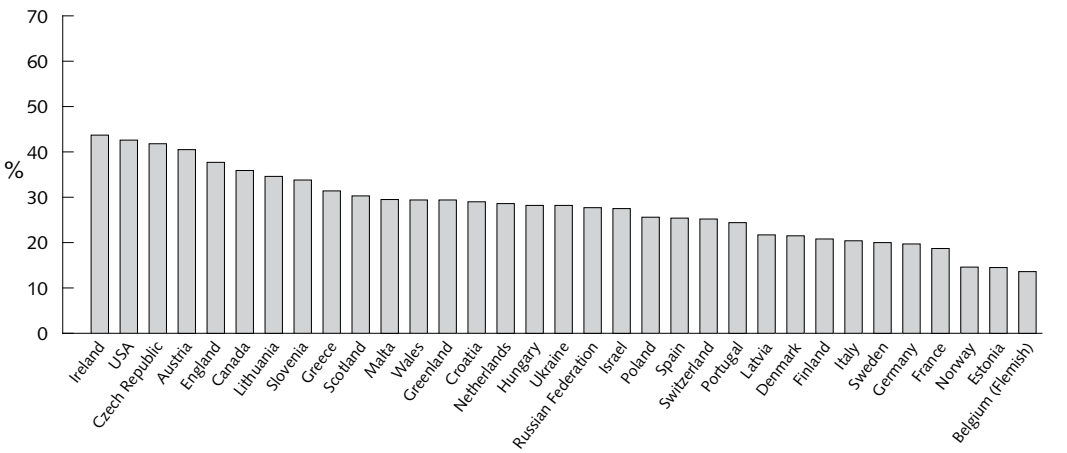

FIG. 2. Proportion of 13-year-old girls meeting recommended levels of physical activity, defined as at least 1 hour of moderate intensity activity on 5 or more days of the week. Data from the 2001/2002 Health Behaviour in School Aged Children Survey [20]

from the 2001/2002 Health Behaviour in School Aged Children Survey [20].

Physical activity is an individual characteristic that is at best moderately stable across childhood into adolescence and adolescence into young adulthood [56–58]. The relatively low to moderate interage correlations for indicators of physical activity across childhood and adolescence into young adulthood reflect, to some extent, the limitations of the methods used to estimate physical activity. The physical activities of children and adolescents are influenced by a variety of factors [10, 59, 60]. Available data are derived largely from samples of European ancestry. Most correlates of activity are aspects of the social environment and highlight a potentially important role for interactions among correlates of physical activity. Notably lacking in the studies are potential correlates of physical activity associated with growth and maturity (with the exception of BMI), physical fitness, and proficiency in movement skills. Relationships with BMI are equivocal [60]. Some data suggest a significant effect of variation in biological maturity status on physical activity. Controlling for individual differences in maturity status (percentage of predicted mature height; see Malina et al. [10]) reduced the sex difference in physical activity in children 5 to 9 years of age; moreover, the relationship between activity and maturity was stronger in males than in females [61]. Activity scores of youths in the

TABLE 4. Recent estimates of physical activity and inactivity in several countries: Global School-Based Student Health Survey, children 13–15 years of age [22]

Area	Year	% Active[a]		% Inactive[b]	
		Boys	Girls	Boys	Girls
Chile, Metropolitan Region	2004	15.4 ± 2.1	7.0 ± 1.9	40.0 ± 4.3	51.3 ± 3.0
Chile, region I	2004	18.6 ± 3.6	8.2 ± 2.1	44.3 ± 5.6	52.7 ± 5.5
Chile, region V	2004	15.3 ± 1.9	7.5 ± 2.0	39.9 ± 3.4	53.1 ± 5.4
Chile, region VIII	2004	12.4 ± 2.8	5.4 ± 2.3	31.9 ± 4.1	48.0 ± 5.9
China, Beijing	2003	24.9 ± 2.8	17.2 ± 3.4	21.3 ± 3.4	22.6 ± 3.1
China, Hangzhou	2003	12.1 ± 2.0	9.0 ± 2.5	22.4 ± 2.5	27.8 ± 5.3
China, Wuhan	2003	18.6 ± 4.2	10.0 ± 2.6	21.8 ± 4.1	16.0 ± 3.2
China, Wurumqi	2003	20.7 ± 2.4	10.7 ± 2.0	20.8 ± 2.4	22.5 ± 3.6
Guyana	2004	16.8 ± 5.0	15.2 ± 4.0	40.3 ± 6.4	32.4 ± 5.1
Jordan	2004	18.9 ± 4.4	13.6 ± 3.2	41.8 ± 5.4	40.5 ± 3.7
Kenya	2003	14.2 ± 1.9	10.1 ± 2.1	40.0 ± 3.9	35.9 ± 3.7
Kenya, Central Region	2003	14.9 ± 3.2	8.8 ± 1.9	48.0 ± 3.9	36.4 ± 3.5
Kenya, Lake Region	2003	13.3 ± 1.6	12.5 ± 3.6	42.3 ± 5.6	37.0 ± 7.9
Kenya, Pastoral Region	2003	12.1 ± 5.0	8.0 ± 7.6	29.3 ± 8.2	34.6 ± 10.2
Namibia	2004	10.2 ± 2.1	9.7 ± 2.2	29.6 ± 2.9	31.8 ± 2.6
Namibia, Central Region	2004	8.4 ± 3.5	8.3 ± 3.4	32.4 ± 4.9	32.5 ± 7.0
Namibia, Northeast Region	2004	11.7 ± 3.8	9.8 ± 3.3	29.3 ± 3.9	27.9 ± 6.2
Namibia, Northwest Region	2004	9.3 ± 4.2	10.9 ± 4.3	27.6 ± 6.0	28.7 ± 3.4
Namibia, Southern Region	2004	11.4 ± 2.7	8.2 ± 2.6	32.0 ± 3.9	39.3 ± 4.0
Oman	2005	32.1 ± 3.6	13.5 ± 2.5	33.2 ± 3.2	35.3 ± 6.6
Philippines	2003	8.5 ± 2.7	6.8 ± 1.6	25.2 ± 5.5	29.9 ± 4.3
Philippines, Luzon	2003	9.6 ± 3.8	7.7 ± 2.1	27.2 ± 7.4	32.7 ± 5.7
Philippines, Mindanao	2003	5.7 ± 2.1	4.0 ± 2.2	19.9 ± 4.0	20.5 ± 6.2
Philippines, Visayas	2003	6.2 ± 1.9	4.3 ± 1.4	20.9 ± 8.8	23.8 ± 4.1
Uganda	2003	16.1 ± 3.8	14.4 ± 2.6	26.7 ± 3.9	27.7 ± 4.7
Uganda, rural	2003	13.6 ± 3.9	11.4 ± 4.2	22.0 ± 5.7	20.9 ± 4.6
Uganda, urban	2003	18.3 ± 6.0	16.9 ± 2.9	30.9 ± 4.3	33.2 ± 7.0
Venezuela, Barinas	2003	11.8 ± 3.3	4.8 ± 1.8	22.8 ± 3.9	18.5 ± 2.5
Venezuela, Lara	2003	14.0 ± 3.2	5.1 ± 1.5	28.6 ± 6.0	31.4 ± 7.6
Zambia	2004	9.7 ± 3.5	10.2 ± 3.4	33.1 ± 4.8	32.7 ± 3.7
Zimbabwe, Bulawayo	2003	14.4 ± 4.5	11.6 ± 2.7	42.9 ± 5.4	42.0 ± 5.6
Zimbabwe, Harare	2003	16.1 ± 3.4	12.9 ± 3.9	42.8 ± 4.2	44.7 ± 5.4
Zimbabwe, Manicaland	2003	13.0 ± 4.5	10.5 ± 3.4	38.9 ± 4.3	42.0 ± 6.8

a. Percentage of students who were physically active on all 7 days for a total of at least 60 minutes per day during the past 7 days.
b. Percentage of students who spent 3 or more hours per day sitting and watching television, playing computer games, talking with friends, or doing other sitting activities during a typical or usual day.

longitudinal Saskatchewan Bone Mineral Accrual Study declined with age from 10 to 18 years, and males were more active than females throughout the study period [62]. However, aligning activity scores on peak height velocity (PHV) independently of chronological age reduced the difference between sexes in activity prior to PHV and eliminated it at and after PHV.

Other potential correlates of physical activity are sport involvement and physical fitness. Boys and girls 12 to 14 years of age participating in organized community youth sports programs expend more energy per unit body mass daily and spend less time watching television than nonparticipants [63]. Among Taiwanese youths 12 to 14 years of age, those classified as fit in the 1-mile (approximately 1,500 m) run and the sit-and-reach test had, on average, greater estimated daily energy expenditure (kcal/kg/day) than those classified as unfit in these tests. In contrast, youths classified as fit and unfit in situps and sum of skinfolds did not differ in estimated energy expenditure [64]. Nevertheless, there is considerable variation and overlap among fit and unfit youths, which highlight earlier observations of Andersen et al. [65] (p. 435) that "among those with the poorest fitness, there are sedentary, moderately active and very active children. Similarly, there are sedentary, moderately active and very active children among those who are in excellent physical condition."

Physical fitness during childhood and adolescence

Physical fitness changes with age, growth, and maturation independently of physical activity. Basic movement skills develop during early childhood and reach mature form by about 5 to 8 years of age. This age interval appears to be a transitional period for fitness tests. In addition to interindividual variation in movement skills, the application of basic skills to specific test situations must be practiced or learned. Mean performance curves of boys and girls in a variety of fitness tasks (grip/arm strength, dash, standing long-jump, vertical jump, shuttle run, and others) show more or less linear improvement from 6 to about 13 to 14 years. Subsequently, the curves for boys indicate an acceleration (adolescent spurts) while those for girls show slight improvement (less intense spurt) in some items and a plateau in others [10]. There is much overlap in performance-related fitness tasks between the sexes during childhood. The mean performances of girls generally fall within one standard deviation of the corresponding performances of boys in early adolescence; subsequently, the mean performances of girls are often outside the limits defined by one standard deviation below the mean performances of boys. Exceptions to the preceding trends are distance-throwing and flexibility. The majority of girls do not show proficiency in throwing, whereas flexibility declines with age during childhood and increases during adolescence and is, on average, greater in girls than in boys at all ages [10].

Indicators of submaximal (PWC_{170}) and maximal ($\dot{V}O_{2max}$, peak $\dot{V}O_2$) cardiorespiratory fitness show similar trends. PWC_{170} (kgm/min) increases, on average, linearly with age between 7 and 17 years; it almost triples in boys and more than doubles in girls over this interval. Similar trends are apparent for power output at other submaximal heart rates (e.g., PWC_{130}, PWC_{150}). $\dot{V}O_{2max}$ (L/min) also increases, on average, from 8 years on. The increase is rather continuous until about 16 years in boys and 13 years in girls. Subsequently, $\dot{V}O_{2max}$ remains at a plateau through adolescence in girls. On average, absolute PWC_{170} and $\dot{V}O_{2max}$ are greater in boys than in girls at all ages [10]. Reliable determinations of cardiorespiratory fitness in younger children are not extensive and are of doubtful validity.

Indicators of motor and aerobic fitness track at moderate levels from childhood into adolescence and through adolescence into young adulthood [56, 57]. They are influenced by changes in body size, physique, and body composition associated with growth and maturation. Size is not ordinarily controlled in tests of performance-related fitness, but body mass is routinely controlled in measures of cardiorespiratory fitness. In contrast to age-associated improvement, submaximal power output remains almost constant from age to age when expressed per unit body mass, i.e., PWC_{170} (kgm/kg/min). Trends are generally similar for $\dot{V}O_{2max}$ per unit body mass (ml O_2/kg min), but there is some variation between longitudinal and cross-sectional studies and with treadmill and cycle ergometer protocols [10]. Comprehensive reviews of cross-sectional and longitudinal data for relative $\dot{V}O_{2max}$ indicate a constant level from 8 to 18 years in boys and a linear decline with age in girls [66, 67]. Recognizing the limitations of correcting for body mass, there is considerable discussion of appropriate scaling factors for $\dot{V}O_{2max}$ and muscular strength [10].

Measures of fitness are influenced by interindividual differences in maturation, particularly during adolescence. Correlational and multivariate analyses indicate significant associations between maturity status and indicators of performance- and health-related fitness [10]. The associations are mediated in part by maturity-associated variation in body size and composition. Within an age group, those advanced in maturation are, on average, taller and heavier, have a larger fat-free mass (especially males) and fat mass (especially females) than those who are delayed in maturation. Maturity-associated variation is most pronounced during adolescence. Early-maturing adolescents of both sexes are absolutely stronger and have greater abso-

lute $\dot{V}O_{2max}$ than those who are late-maturing. After adjustment for variation in body mass, the differences are negligible or reversed. Corresponding data for performance-related fitness variables (jumps, dashes, etc.) are more variable. Early-maturing boys tend to perform better than late-maturing boys, but the differences between early- and late-maturing girls, on average, are quite small. It is thus possible that a youngster could be classified as "fit" or "unfit" due largely to his or her maturity status.

Although size, physique, body composition, and maturity status account for substantial portions of variation in fitness, considerable amounts of variation are not accounted for by these variables. Performances on fitness tests are influenced by motivational factors, opportunity for practice, exposure to appropriate instruction, and perhaps other factors in the cultural environment. Habitual physical activity is an additional factor.

Physical activity, growth, and maturation

The individual adapts to the stresses imposed by habitual physical activity. The responses are, to a large extent, not sufficient to significantly alter growth in height and indicators of skeletal, sexual, and somatic maturation. Physical activity can be important in the regulation of body weight, fatness, and the structural and functional integrity of bone and skeletal muscle tissues [10].

Physical activity is presumably important, but it is not known how much activity is necessary to support growth and maturation. The day-to-day activities of childhood and adolescence are apparently adequate to maintain the integrity of growth and maturation processes, with the possible exception of adipose tissue. Physical inactivity and excessive energy intake are associated with greater levels of fatness. Both are major contributors to the current epidemic of obesity in children and adolescents.

Physical activity and physical fitness

It is often assumed that physical activity is related to physical fitness, i.e., the more habitually active are more fit, and that the relationship between activity and fitness is causal. This is not necessarily so. Relationships between regular physical activity and indicators of physical fitness are generally low to moderate in children [68, 69] and adolescents [70–72]. Physical activity accounts for a relatively small percentage of the variation in some indicators of fitness [64, 73, 74]. Several factors probably contribute to these observations in children and adolescents: measures of activity are imperfect and tests of fitness vary in validity and reliability, and levels of habitual physical activity among youths do not regularly reach elevated aerobic levels for sustained periods of time. Proficiency in movement skills may also be a factor. Some evidence suggests that children 5 to 6 years of age and adolescents 13 to 15 years of age who are proficient in movement skills have higher levels of cardiorespiratory endurance [75, 76].

Components of fitness change with growth and maturation independent of physical activity, and it is difficult to partition effects of activity from the expected changes. For example, a late-maturing boy might not do well on fitness tests requiring muscular strength, simply because he is smaller than average and has less muscle mass. Similarly, an early-maturing boy might have an advantage on some fitness tests because he is taller, heavier, and stronger than average-maturing boys. Early-maturing girls might not do well on fitness tests requiring support or projection of the body because they tend to be fatter than average-maturing and late-maturing girls.

Time spent in inactive pursuits such as television viewing, video games, and other sedentary activities is another factor that may influence physical fitness. The relationship between time spent viewing television and indicators of health-related fitness is not strong [77], but some evidence increasingly suggests a does–response relationship between television viewing and indicators of adiposity, specifically obesity [78, 79].

It is possible that the relationship between physical activity and fitness is masked, in part, by the normal range of variability in heterogeneous samples of children and adolescents. The relationship may be more apparent in comparisons of groups at the extremes of the physical activity continuum. Although methods of classifying youths as active and inactive vary among studies, both cross-sectional [64, 80] and longitudinal [81–83] observations indicate that more active youths are more fit in cardiorespiratory endurance tasks. Corresponding comparisons for other components of physical fitness are inconsistent, suggesting that the effect of habitual physical activity may be specific to cardiorespiratory endurance. Moreover, physical activity is only one of several factors that influence physical fitness.

Physical activity, fitness, and undernutrition

The growth status, physical activity, and fitness of children and adolescents resident in many developing areas of the world are compromised by chronic undernutrition and associated health problems. The quantification of energy expenditure of schoolchildren living under marginal nutritional circumstances has been a primary focus of study rather than the quality of physical activity and movement experiences relevant

to the development of proficiency of movement skills and physical fitness. School-aged children of marginal to poor nutritional status have decreased total daily energy expenditure and energy expended in physical activity. This is related in part to smaller body size. Reduced activity may limit practice of movement skills in play and games and in turn contribute to performance deficiencies. Quasi-experimental observations in Colombia suggest that mild-to-moderately undernourished boys differ from boys with adequate nutrition in their capacity to increase energy expenditure in physical activity when given the opportunity to participate in a sports program [84, 85]. The undernourished boys simply could not keep up with the better-nourished boys during sport activities.

The small body size and reduced muscle mass associated with chronic, mild-to-moderate undernutrition influences the strength, performance, PWC_{170}, and $\dot{V}O_{2max}$ in samples of school-aged children in developing areas of the world. The absolutely lower levels of fitness are generally proportional to the reduced stature and body mass of children and adolescents living under conditions of chronically marginal nutritional circumstances [84–91]. When adjusted for body mass, the differences are reduced to negligible levels and in some cases reversed, although there are exceptions [86]. The effect of small body size on performances of schoolchildren varies among motor tasks and between populations [89, 90, 92], and the significance of the BMI for the strength and motor fitness of children with a history of chronic undernutrition is different from that of well-nourished children [93].

Infectious load associated with intestinal parasites may influence physical activity and fitness of children with chronic undernutrition. Treatment of undernourished school-aged Kenyan children for hookworm, whipworm (trichuris), and roundworm (ascaris), for example, was associated with increased levels of spontaneous physical activity, improved cardiovascular fitness (step test), and improved growth and appetite [94, 95]. Similarly, treatment of undernourished school-aged children for schistosomiasis was also associated with improved step-test performance [96]. A related factor is illness, and it is well known that during periods of illness, children tend to be less active. Among mild-to-moderately undernourished school-aged Kenyan children, moderate physical activity on the playground was negatively correlated with the frequency of mild illness, especially respiratory and gastrointestinal illness. On the other hand, light activity on the playground was positively correlated with the frequency of illness [97].

Since the performances of schoolchildren living under conditions of mild-to-moderate undernutrition are largely proportional to their reduced body size, these children are sometimes described as "small but efficient." The same applies to adults who were raised under impoverished health and nutritional conditions, which has prompted the "small but healthy" hypothesis [98, 99]. Under this hypothesis, the smallness (growth stunting) and reduced muscle mass are the primary adaptations to conditions of chronic undernutrition, and capacity for physical work is not impaired. However, the conditions that produce chronic undernutrition, i.e., inadequate diet, poor home environments, infectious and parasitic diseases, and so on, are, themselves, unhealthy [100]. Stunted growth and reduced physical fitness in a variety of tasks are the results of these impoverished and unhealthy conditions. In the real world of children, adolescents, and adults, however, physical tasks and corresponding performances are not scaled for individual differences in body size!

Physical activity, fitness, and obesity

The literature on physical activity and energy expenditure in childhood and adolescent obesity is equivocal [101–104]. Absolute energy expenditure (measured by doubly labeled water) is greater in the obese, but after "normalization" for differences in body composition, energy expenditure is the same in the obese and the nonobese [105]. Data based on heart-rate recordings suggest lower activity levels in free-living obese adolescents [106], and activity data based on questionnaires or time and motion analyses suggest that obese children and adolescents are less active than their lean peers [107, 108]. Among children 3 to 5 years of age, observational and accelerometry data indicate lower activity during the preschool day in the overweight than in the nonoverweight [109].

The physical fitness of obese children and adolescents is compromised on tasks that require movement or projection of the body through space, such as runs and jumps. From a mechanical perspective, excess fat represents an inert load (dead weight) that must be moved. In a national sample of Flemish girls 7 to 17 years of age, the obese (the fattest 5% in each age group) attained, on average, poorer results than lean girls (the leanest 5% in each age group) in the shuttle run (speed and agility), standing long-jump and vertical jump (power), flexed arm hang (functional strength), situps and leg lifts (abdominal strength), and flamingo stand (balance). The obese and lean did not differ in speed of limb movement (plate tapping) and the sit-and-reach test (flexibility). Obese girls were absolutely stronger (static arm pull strength) and generated more absolute power on a cycle ergometer (PWC_{170}), but the results were reversed after adjustment for differences in body mass [110]. Corresponding data for obese and nonobese Belgian boys [111] and boys and girls [112] on the same tests showed similar results. Obese boys 9 to 13 years of age were absolutely stronger than the nonobese in isometric and isokinetic strength tests, but the groups did not differ in muscle contractile

characteristics or in the rate of motor unit recruitment [113, 114]. Reduced cardiovascular fitness in the obese is evident in submaximal ergometry [115], maximal aerobic power per unit body mass [116–118], heart-rate response to a step test [110, 116], and endurance time on a maximal treadmill test [119].

Secular changes in physical activity and fitness

There are few data on secular trends in physical activity among children and youths. This is due to several factors. First, there has historically been more emphasis on the measurement of physical fitness in youths; in contrast, large-scale surveys of physical activity in children and adolescents date only to the 1980s (see above). Second, health risks associated with physical inactivity in children have received attention only recently, due in part to the epidemic of obesity. And third, standardized methodologies for the collection of physical activity data over time and adequate physical activity surveillance systems have been lacking in most countries.

General demographic information suggests a decline in habitual physical activity over the past 40 years or so. For example, the number of cars per household and hours per week of television viewing has increased linearly since the mid-1960s in the United Kingdom [120]. Between 1947 and 1999, the estimated percentage of Americans walking or bicycling to work has declined from 26% to 4%, the percentage taking a bus or subway has declined from 25% to 3%, while the percentage driving a vehicle to work has increased from 32% to 87% [121]. Physically active transport to work (or to a bus or subway station) has thus declined systematically, while the potential for physical inactivity has increased over time.

Introduction of television into communities that previously did not have access resulted in a reduction in sports participation [122, 123]. The introduction of television thus displaced physically active leisure pursuits, and the trends suggest a greater impact on youths than on adults. Increased academic demands and accessibility of computers and the internet may also contribute to more sedentary time among youths. For example, the time spent on homework by 6- to 9-year olds in the United States increased from 44 minutes per week in 1981 to more than 2 hours in 1997; the corresponding estimates for 9- to 11-year-olds were 2 hours and 49 minutes in 1981 and more than 3.5 hours per week in 1997 [124]. About 44% of children 2 to 12 years of age and 75% of teenagers were predicted to be on-line by 2002 [125].

An analysis of activities in the daily lives of American children 3 to 12 years of age, based on two surveys in 1981 and 1997, suggested the following trends, which have implications for physical activity: free or discretionary time declined, time in school and day care increased, time viewing television declined, time spent reading or studying at home increased, time in organized activities (hobbies, sports, arts) increased, and time in unstructured activities decreased [126]. The sequel to this review indicates little change over the past few decades in physically active transportation (an average of 8 minutes in 2001), no trend for enrollment in physical education over the past decade among high-school students, and an increase in time spent in organized sports and outdoor activities by younger children between 1981 and 1997 of 73 minutes per week [161].

The trends would seem to suggest an increase in opportunities for sedentary behaviors in contrast to a single sedentary behavior. Television viewing per se may not be a good marker of sedentary behavior. It is more likely that there has been an increase in the number of sedentary behaviors over the past generation—playing video games, personal computer activities, watching DVDs, homework, extracurricular classes (tutoring, art, music), motorized transport to school and other organized activities, and probably others.

Short-term physical activity trends for youths are inconsistent. Canadian youths 12 to 19 years of age indicated an increase in self-reported leisure-time physical activity between 1981 and 1988, but there was no change between 1988 and 1998 [127]. The majority of Canadian youths do not meet the recommendations for daily physical activity energy expenditure. The percentage of American high-school students who attended physical education classes daily declined from 42% in 1991 to 28% in 2003 [128]. On the other hand, the percentage of those enrolled in physical education classes who were physically active for more than 20 minutes in a class on 3 to 5 days per week increased only slightly between 1991 and 2003 (from 32% to 39%). The percentage of high-school students reporting participation in physical activities that made them sweat and breathe harder for more than 20 minutes on 3 of the past 7 days also changed only slightly between 1993 (66%) and 2003 (63%) [14].

Measures of physical fitness are, in part, related to body size and maturity status. Hence, the increase in size and acceleration in maturity that characterize the secular trend have implications for physical fitness [10]. Measures of fitness are also influenced by changes in lifestyle, specifically reduced physical activity, which may interact with changes in body size.

Interpretation of secular changes in static strength must be tempered with caution, because it is likely that measuring instruments have changed over time. Motivation is an additional factor in obtaining maximal efforts. Allowing for these caveats, secular gains are apparent in height, weight, and grip strength in Belgian (1830s to 1971), American (1899 to 1964 and 1934–35 to 1958–59) and Japanese (1923 to 1969) children and adolescents. Although there is some variation, secular gains in strength appear to be proportional to changes

in height and weight [129]. Data on other measures of static strength are less extensive. Relative to secular gains in height, the back strength of Japanese children and adolescents in 1969 was proportionally less than in 1929. Back strength is a more difficult measure to obtain than grip strength and may be influenced by minor variations in technique. Pushing and pulling strength of the shoulders in 13-year-old American youths showed variable changes over an interval of 24 years (1934–35 to 1958–59) [129].

Some recent data for European youths suggest declines in muscular strength. Data for eight measures of strength were presented for Danish youths of the same height (150 cm) in 1956 and 1981; boys and girls in 1981 were not as strong as peers of the same size in 1956 [130]. Russian youths 11 to 17 years of age in the mid-late 1980s were not as strong (as measured by grip) as peers in the 1960s. Data for height and weight over this interval are apparently not available, but mean ages at menarche were stable from the early 1960s to the early 1980s [131].

Performance-related fitness is more difficult to evaluate over time, since tests and conditions of testing are variable. One of the few test items that is reasonably consistent in test batteries over time and whose testing protocol is generally described in a standard manner is the standing long jump, an indicator of power and coordination. Comparisons of American boys and girls 11 to 15 years of age between the mid-1920s [132] and 1958 through 1985 [35–37] indicate little evidence of secular improvement in mean jumping performance between the mid-1920s and 1958, improvement from 1958 to 1965 in all age groups, and little change from 1965 through 1985. The major improvement in performance from 1958 to 1965 reflected national emphasis on the physical fitness of American youths and fitness testing in schools, and to some extent practice effects. After 1965, national emphasis on the performance aspects of physical fitness declined [133].

Heights and weights of the early and more recent samples of American youths are not available for evaluation of the relationship between secular change in size and performance between the 1920s and the 1960s. However, the secular improvement in the jumping performance of Czechoslovak boys 11 to 15 years of age was essentially a function of the secular increase in body size [129]. More recent data from 11- to 15-year old Czech youths indicate continued secular increases in height, weight, and performance on the standing long jump from 1966 to 1987 [134], but it was not possible (given the manner of reporting) to relate gains in jumping performance to changes in body size.

Secular data for other performance-related fitness tests are more variable. Among Japanese children measured (height) and tested in 1929 (standing long jump), 1935 (100 m dash, boys) and 1939 (50 m dash, girls), and 1969 (with the same tests), boys and girls in 1969 did not perform in the jump as well as their peers in 1929, and there were virtually no differences in performance on the dashes between 1935–39 and 1969 [129]. Relative to height, which increased over time, children in the more recent samples did not perform as well.

Beginning in the 1960s, physical fitness test batteries were commonly administered to school children in many countries. The batteries included a variety of performance-related items, such as dashes, jumps, shuttle runs, ball throws for distance, situps, distance runs, etc. Comparisons of four national surveys in the United States between 1958 and 1985 indicated major improvements in the fitness of 10- to 17-year old youths of both sexes between 1958 and 1965, but there was little change in fitness from 1965 to 1985 [35–37]. As noted, the improvement in fitness from 1958 to 1965 reflected in part the national emphasis on physical fitness testing in schools in the 1960s [133]. A contributing factor to the national emphasis on fitness was the poor performance-related fitness of American youths in 1958 compared with British youths in 1958–59 [135]. The only fitness item in which American boys surpassed British boys was the softball throw for distance, while American and British girls did not differ in this test item. Interest in performance-related fitness has declined since the 1960s as focus has shifted to health-related physical fitness [133]. The heights and weights of American children, on average, did not change appreciably between the 1960s and the 1980s [10]. The results of the Chrysler AAU Physical Fitness Program, which largely used health-related fitness items, indicated only small changes in the fitness of 6- to 17-year-old American youths in the decade of the 1980s, but performance on endurance runs declined [136]. Comparisons of reference values (norms) for tests of health-related fitness of 10- to 17-year-old American youths in the late 1970s [6] and in 1984 [12] indicate generally poorer performances by youths in 1984 in 1-mile run/walk times and number of situps, and variable results for the sit-and-reach test [137]. The results may reflect sampling (convenience versus probability) and testing procedures (volunteer teachers versus trained field staff and teachers). More recent national fitness data are not available for American youths.

Height increased systematically across surveys at 10-year intervals from 1965 to 1995 in children from southwestern Poland, but the same was not true for performance items [138]. With the 1965 data as the reference, performances in strength and endurance tasks declined in each 10-year survey so that the values for the 1995 survey were lowest in comparison with the 1965 values. Declines in strength and endurance were most marked after the ages of 12 to 14 years in both sexes. On the other hand, running speed did not change between 1965 and 1985, but declined in the 1995 survey. Agility did not change appreciably among

the four surveys. More recently, comparisons of 7- to 19-year-old Polish youths in 1979, 1989, and 1999 on the EUROFIT test indicate improvement between 1979 and 1989 but deterioration between 1989 and 1999 in both sexes [44]. The regression in fitness was especially apparent in power of the upper and lower extremities, running speed (sprints), and endurance distance run.

Five tests of fitness were administered to 16-year-old Swedish secondary school children in 1974 and 1995: two-hand lift, vertical jump, situps, bench press, and 9-minute run [139]. Both boys and girls performed more poorly on the bench press, situps, and 9-minute run in 1995 than 1974. Performances on the vertical jump improved over time in boys but not in girls, whereas both boys and girls improved in the two-hand lift. The increase in BMI over time explained variable portions of the variance in the fitness tests, leading the authors to suggest a significant role for decreased physical activity.

Comparisons of South Australian 10- to 11-year-old youths in 1985 and 1997 indicated, on average, a decline in aerobic fitness (1.6 km run/walk time) and running speed (50-m dash), but no differences in the standing long jump [140].

The pattern of secular change in the physical fitness and motor ability of 12- to 17-year old Japanese school youths between 1964 and 1997 shows a somewhat variable pattern [141, 142]. Composite fitness and motor ability (performance) scores increased from 1964 to 1974, increased slightly (fitness) or remained stable (motor) between 1975 and 1985, and then declined from 1986 through 1997. The early improvements in fitness and motor ability may be related in part to secular gains in height from 1964 to 1984, about 5 cm in boys and 3 cm in girls at 17 years of age [141], whereas the subsequent declines in fitness may be related to changing patterns of physical activity [142]. A different analysis of essentially the same data attributed improvements (1964–74) to national emphasis on practice and fitness in the schools, the leveling-off of fitness (1975–85) to increased television viewing in Japanese homes, and subsequent declines in fitness to greater emphasis on scholarship and activity for pleasure in schools and increased use of video games [143].

Several surveys summarized above indicate a decline in aerobic fitness as assessed by field measures of endurance, primarily distance runs. Evaluation of performances of 12- to 15-year old Australian youths [144] and 6- to 19-year-old youths from 11 countries [145] on the 20-m multistage shuttle run test suggests a systematic decline in aerobic fitness between 1981 and 2000. The overall mean decline weighted for sample size in the international comparison was –0.43% per year.

Data for maximal aerobic power ($\dot{V}O_{2max}$) for American youths are available back to the late 1930s in boys and the 1960s in girls; hence, insights into secular changes are possible [146]. However, most studies are based on rather small samples combined across several age groups, and it is likely that overweight children were not included, given the maximal exercise protocols. Cycle data were adjusted for differences compared with the treadmill protocol. For absolute aerobic power (L/min), the regression is flat from the late 1930s to 2000 in boys 6 to 12 years of age, but that for boys 13 to 18 years of age suggests an increase over time. The corresponding regression line is stable from the 1970s in girls 6 to 11 years of age and from the 1960s in girls 12 to 14 years of age, but among older girls 15 to 18 years of age it is curvilinear, suggesting an increase from the early 1960s to the late 1970s and a decline into the late 1990s. When the data are expressed per unit body weight (ml/kg/min), accommodating, to some extent, secular change in body mass, the regression lines indicate fairly stable levels of relative maximal aerobic power between the 1930s and the present in boys; the trends for relative maximal aerobic power in girls are similar to those for absolute values.

An important component of health-related physical fitness is adiposity, viewed most often in terms of the BMI. The prevalence of obesity among children and adolescents has increased dramatically since the mid-1970s in many countries throughout the world [147, 148]. Data from Canada and the United States show increases in both the mean BMI and the prevalence of overweight and obesity in children and adolescents [149–153]. However, within each age group, there appears to have been a greater increase in the upper percentiles of the BMI, producing an effect of increasing skewness in the distribution over time [152]. Trends in some countries suggest an increase in the prevalence of obesity without corresponding increases in the median BMI. Among Belgian boys 12 to 18 years of age, for example, there were no changes in the median BMI between 1969–74 and 1990–93, but increases in the 85th percentile value ranged from 0.5 to 1.3 kg/m^2 and those in the 95th percentile ranged from 0.9 to 1.9 kg/m^2. Corresponding data for females indicated a small increase in median BMIs between surveys, but larger increases in the 85th (1.0 to 2.2 kg/m^2) and 95th (1.1 to 2.7 kg/m^2) percentile values [154].

Obesity is a major factor in the metabolic syndrome, which is characterized by a clustering of risk factors associated with chronic disease. Proposed definitions of the metabolic syndrome for adolescents include indicators of abdominal obesity, elevated triglycerides, high blood pressure, elevated fasting glucose, and low HDL cholesterol [155, 156]. A significant percentage of children and youths have features of the metabolic syndrome, and the risk of metabolic syndrome is greater in overweight and obese children and youths [155, 156]. Few studies have evaluated the impact of physical activity on the metabolic syndrome in youths [157], but several features of the syndrome cluster with indicators

of physical activity and inactivity and cardiorespiratory fitness [158, 159].

Rationale for including measures of physical activity and physical fitness

It is generally assumed that children and adolescents need regular physical activity to support normal growth and maturation, but an optimal amount of daily activity has not been specified. The apparent decline in physical activity and physical fitness over the past generation or two, the recent increase in the prevalence of obesity among youths worldwide, the emergence of risk factors for several adult diseases (metabolic and cardiovascular) during childhood and adolescence, and the potential role of physical activity and physical fitness in reducing the risk of disease would seem to imply a need to include a measure of physical activity and physical fitness in the development of an International Growth Standard for Preadolescent and Adolescent Children.

A questionnaire for habitual physical activity for youths 10 years of age and older should be adequate. The validity and reliability of an instrument that is relevant cross-culturally would need to be established, especially since a potential international growth standard will probably be utilized at the individual level.

Inclusion of measures of physical fitness is a more complex issue. Cardiorespiratory fitness is the component of physical fitness most relevant to physical activity and health and should be included in developing the standard. An endurance shuttle run should be adequate for this purpose. The international standard will already include the BMI (a component of health-related fitness).

Given the importance of muscular strength and endurance to undertaking daily tasks in many areas of the world, a measure of muscle function should be included in the development of the standard. Timed curl-ups and the flexed (bent) arm hang provide an indication of abdominal and upper body muscular strength and endurance, respectively.

Allowing for individual differences in physical activity and physical fitness, the development of growth standard values for children and adolescents in different subgroups of activity and fitness should perhaps be considered. To this end, specific components of an established test battery, such as EUROFIT [33], should be appropriate.

Feasibility of including measures of physical activity and physical fitness

Incorporation of an internationally valid and reliable questionnaire to estimate habitual physical activity among those 10 years of age and older is very feasible. Evidence suggests that a record of 7 days provides a reliable estimate of the usual pattern of physical activity in youths in grades 4 through 12 [160].

Incorporation of field measures of cardiorespiratory fitness and muscular strength and endurance is more complicated because of administrative issues (facilities, trained staff, etc.) and potential medical or health complications for some youths free from overt disease.

Summary

Physical activity is a "behavior," whereas physical fitness is a "state." Both change with growth and maturation. Physical activity is presumably important, but it is not known how much activity is necessary to support growth and maturation.

Energy expenditure (EE) associated with physical activity is the most variable component of total daily energy expenditure (TEE). The ratio of TEE to resting EE (REE) provides an estimate of "physical activity level" (PAL). The PAL increases with age during childhood and adolescence, but is particularly low in sedentary children.

More physical activity data are available for children and adolescents 10 years of age and older than for younger children. The data are based largely on questionnaires and interviews. Activity level appears to peak at about 12 to 14 years of age and then declines across adolescence, but the timing and magnitude of the decline varies among studies and with measuring instrument.

Relationships between regular physical activity and indicators of physical fitness are generally low to moderate in children and adolescents. Components of fitness change with growth and maturation independent of physical activity, and it is difficult to partition effects of activity from the expected changes.

School-aged children of marginal to poor nutritional status show decreased total daily energy expenditure and energy expended in physical activity. The small body size and reduced muscle mass associated with chronic, mild-to-moderate undernutrition negatively influence activity and many indicators of fitness. The hypothesis of "small but efficient" is erroneous. Stunted growth and reduced levels of physical fitness are often the results of impoverished and unhealthy conditions.

The evidence on physical activity and energy expenditure in childhood and adolescent obesity is equivocal, while the physical fitness of obese children and adolescents is compromised on tasks that require movement or projection of the body through space, e.g., runs and jumps.

General demographic data suggest a decline in the amount of habitual physical activity over the past 40 years or so. Although the increase in size and acceleration in maturity that characterize the secular trend have implications for physical fitness, measures of fitness are also influenced by changes in lifestyle, specifically reduced physical activity. Several surveys indicate a secular decline in indicators of physical fitness.

References

1. Tremblay MS, Willms JD. Is the Canadian childhood obesity epidemic related to physical inactivity? Int J Obes Relat Metab Disord 2003; 27:1100–5.
2. US Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. Atlanta, Ga, USA: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.
3. National Institutes of Health. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults—the evidence report. Obes Res 1998;6(suppl 2):51S–209S (published erratum appears in Obes Res 1998;6:464).
4. Katzmarzyk PT, Janssen I, Ardern CI. Physical inactivity, excess adiposity and premature mortality. Obes Rev 2003;4;257–90.
5. Clarke HH. Basic understanding of physical fitness. Physical Fitness Research Digest 1971; series 1, no. 1.
6. American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD). Health related physical fitness test manual. Reston, Va, USA: AAHPERD, 1980.
7. American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD). Technical manual, health related physical fitness. Reston, Va, USA: AAHPERD, 1984.
8. Fitness Canada. Fitness and lifestyle in Canada. Ottawa: Government of Canada, 1983.
9. Bouchard C, Shephard RJ. Physical activity, fitness, and health: the model and key concepts. In: Bouchard C, Shephard RJ, Stephens T, eds. Physical activity, fitness, and health. Champaign, Ill, USA: Human Kinetics, 1994: 77–88.
10. Malina RM, Bouchard C, Bar-Or O. Growth, maturation, and physical activity, 2nd ed. Champaign, IL, USA: Human Kinetics, 2004.
11. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport 2000;71(2 suppl):S59–73.
12. Ross JG, Gilbert GG. Summary of findings from National Children and Youth Fitness Study. J Phys Educ Rec Dance 1985;56(Jan):1–48.
13. Ross JG, Pate RR. Summary of findings from National Children and Youth Fitness Study II. J Phys Educ Rec Dance 1987;58(Nov–Dec):49–96.
14. Centers for Disease Control. YRBSS: Youth online comprehensive results. 2005. Available at: http://www.cdc.gov/HealthyYouth/index.htm. Accessed 13 August 2006.
15. Stephens T, Craig CL. The well-being of Canadians: highlights of the 1988 Campbell's Survey. Ottawa: Canadian Fitness and Lifestyle Research Institute, 1990.
16. King AJC, Coles B. The health of Canada's youth: views and behaviours of 11-, 13- and 15-year-olds from 11 countries. Ottawa: Minister of National Health and Welfare, 1992.
17. Chen M, Asami T. Chino-Japanese cooperative study on physical fitness of children and youth, I. Tokyo: Sports Science Committee, Japan Amateur Sports Association, 1986 (in Japanese).
18. Chen M, Asami T. Chino-Japanese cooperative study on physical fitness of children and youth, II. Tokyo: Sports Science Committee, Japan Amateur Sports Association, 1987 (in Japanese).
19. Booth M, Macaskill P, McLellan L, Phongsavan P, Okely T, Patterson J, Wright J, Bauman A, Baur L. NSW schools fitness and physical activity survey 1997. Ryde, New South Wales, Australia: NSW Department of Education and Training, 1997.
20. Currie C, Roberts C, Morgan A, Smith R, Settertobulte W, Samdal O, Rasmussen VB, eds. Young people's health in context: international report from the HBSC 2001/02 survey. WHO Policy Series: Health policy for children and adolescents. Issue 4. Copenhagen, Denmark: WHO Regional Office for Europe, 2004.
21. Statistics Canada. Canadian Community Health Survey, 2004. available at: http://stcwww.statcan.ca/english/sdds/3226.htm. Accessed 13 August 2006.
22. Centers for Disease Control and Prevention. GSHS: Global school-based student health survey. 2006. Available at: http://www.cdc.gov/gshs/index.htm. Accessed 13 August 2006.
23. Engstrom LM. Physical activity of children and youth. Acta Paediatr Scand Suppl 1980;283:101–5.
24. Ilmarinen J, Rutenfranz J. Longitudinal studies of the changes in habitual physical activity of school children and working adolescents. In: Berg K, Eriksson BO, eds Children and exercise IX. Baltimore, Md, USA: University Park Press, 1980:149–59.
25. Renson R, Beunen G, Ostyn M, Simons J, Uytterbrouck J, Van Gerven D, Vanreusel B. Differentiation of physical fitness in function of sport participation. Hermes 1981;15:435–44.
26. Kemper HC, Dekker HJ, Ootjers MG, Post B, Snel J, Splinter PG, Storm-van Essen L, Verschuur R. Growth and health of teenagers in the Netherlands: survey of multidisciplinary longitudinal studies and comparison to recent results of a Dutch study. Int J Sports Med 1983;4:202–14.
27. American Association for Health, Physical Education, and Recreation (AAHPER). Youth fitness test manual. Reston, Va, USA: AAHPER, 1958.
28. American Association for Health, Physical Education, and Recreation (AAHPER). Youth fitness test manual, revised edition. Reston, Va, USA: AAHPER, 1976.
29. Morrow JR, Falls HB, Kohl HW, eds. The Prudential Fitnessgram: technical reference manual. Dallas, Tex, USA: Cooper Institute for Aerobic Research, 1994.
30. Cooper Institute for Aerobics Research. The Prudential Fitnessgram test administration manual. Dallas, Tex, USA: Cooper Institute for Aerobics Research, 1992.
31. Canadian Association for Health, Physical Education and Recreation (CAHPER). Fitness performance test manual for boys and girls 7 to 17 years of age. Ottawa: CAHPER, 1966.
32. Canadian Association for Health, Physical Education and Recreation (CAHPER). The physical working capacity of Canadian children aged 7 to 17. Ottawa: CAHPER, 1968.
33. Council of Europe, Committee for the Development of Sport. EUROFIT: European test of physical fitness.

Rome: Committee for the Development of Sport within the Council of Europe, 1988.
34. To C-Y. Physical fitness of children in Hong Kong. Hong Kong: Chinese University of Hong Kong, School of Education, 1985.
35. Hunsicker PA, Reiff GG. A survey and comparison of youth fitness 1958–1965. J Health Phys Educ Rec 1966; 37:23–5.
36. Hunsicker PA, Reiff GG. Youth fitness report: 1958–1965–1975. J Phys Educ Rec 1977;48:31–3.
37. Reiff GG, Dixon WR, Jacoby D, Ye GX, Spain CG, Hunsicker PA. The President's Council on Physical Fitness and Sports 1985 National School Population Fitness Survey. Ann Arbor, Mich, USA: University of Michigan Press, 1986.
38. Gauthier R, Massicotte D, Hermiston R, MacNab R. The physical work capacity of Canadian children, aged 7 to 17 in 1983. A comparison with 1968. CAHPER Journal/Revue de l'ACSEPR 1983;50(Nov–Dec):4–9.
39. Simons J, Beunen G, Ostyn M, Renson R, Swalus P, Van Gerven D, Willems E. Construction d'une batterie de tests d'aptitude motrice pour garçons de 12 à 19 ans, par la méthode de l'analyse factorielle. Kinanthropologie 1969;1:323–62.
40. Beunen G, Ostyn M, Simons J, Renson R, Van Gerven D. Chronological and biological age as related to physical fitness in boys 12 to 19 years. Ann Hum Biol 1981; 8:321–31.
41. Beunen GP, Malina RM, Van't Hof MA, Simon J, Ostyn M, Renson R, Van Gerven D. Adolescent growth and motor performance: a longitudinal study of Belgian boys. Champaign, Ill, USA: Human Kinetics, 1988.
42. Pyke JE. Australian health and fitness survey 1985. Parkside, South Australia: Australian Council for Health, Physical Education and Recreation, 1985.
43. Simons J, Beunen GP, Renson R, Claessens ALM, Vanreusel B, Lefevre JAV. Growth and fitness of Flemish girls: the Leuven Growth Study. Champaign, Ill, USA: Human Kinetics, 1990.
44. Przewęda R, Dobosz J. Kondycja fizyczna Polskiej młodzieży (physical condition of Polish youth). Studia i Monografie, Akademia Wychowania Fizycznego Józefa Piłsudskiego w Warszawie, No. 98, 2003 (in Polish).
45. Alexander P. Aptitud fisica, caracteristicas morfologicas, composicion corproral: pruebas estandarizadas en Venezuela de 7.5 a 18.4 años. Caracas, Venezuela: Instituto Nacional de Deportes, 1995.
46. Alexander P, Mota D, Arévalo A. Normas para la evaluacion de la aptitud fisica y caracteristicas morfologicas del estudiante Guatemalteco de 7.5 a 18.4 años. Ciencias de la Actividad Fisica 1993;2:1–54.
47. Nieto GJ, Ordoñez Sanchez ON. Aptitud fisica: pruebas estandarizadas en Colombia. Bogota, Colombia: Instituto Colombiano de la Juventud y el Deporte "Coldeportes," 1994.
48. Barbieri CO. Programa de evaluacion diagnostico e investigacion de la aptitud fisica y la salud. Buenos Aires, Argentina: Instituto Bonaerense del Deporte, 1997.
49. Torun B, Davies PS, Livingstone MB, Paolisso M, Sackett R, Spurr GB. Energy requirements and dietary energy recommendations for children and adolescents 1 to 18 years old. Eur J Clin Nutr 1996;50:S37–81
50. Saris WHM, Elvers JWH, van't Hof MA, Binkhorst RA. Changes in physical activity of children aged 6 to 12 years. In: Rutenfranz J, Mocellin R, Klimt F, eds. Children and Exercise XII. Champaign, Ill, USA: Human Kinetics, 1986:121–30.
51. Verschuur R, Kemper HCG. Habitual physical activity in Dutch teenagers measured by heart rate. In: Binkhorst RA, Kemper HCG, Saris WHM, eds. Children and exercise XI. Champaign, Ill, USA: Human Kinetics, 1985: 194–202.
52. Butte NF, Wong WW, Hopkinson JM, Heinz CJ, Mehta NR, Smith EO. Energy requirements derived from total energy expenditure and energy deposition during the first 2 y of life. Am J Clin Nutr 2000;72:1558–69.
53. Bailey RC, Olson J, Pepper SL, Porszasz J, Barstow TJ, Cooper DM. The level and tempo of children's physical activities: an observational study. Med Sci Sports Exerc 1995;27:1033–41.
54. Caspersen CJ, Pereira MA, Curran KM. Changes in physical activity patterns in the United States, by sex and cross-sectional age. Med Sci Sports Exerc 2000;32: 1601–9.
55. van Mechelen W, Twisk JW, Post BG, Snel J, Kemper HC. Physical activity of young people: the Amsterdam Longitudinal Growth and Health Study. Med Sci Sports Exerc 2000;32:1610–6.
56. Malina RM. Tracking of physical activity and physical fitness across the lifespan. Res Q Exerc Sport 1996;67(3 suppl):S48–57.
57. Malina RM. Physical activity and fitness: pathways from childhood to adulthood. Am J Hum Biol 2001;13: 162–72.
58. Malina RM. Tracking of physical activity across the lifespan. Research Digest, President's Council on Physical Fitness and Sports 2001, Series 3, 14:1–8.
59. Kohl HW, Hobbs KE. Development of physical activity behaviors among children and adolescents. Pediatrics 1998;101(3 pt 2):549–54.
60. Sallis JF, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. Med Sci Sports Exerc 2000;32:963–75.
61. Eaton WO, Yu AP. Are sex differences in child motor activity level a function of sex differences in maturational status? Child Dev 1989;60:1005–11.
62. Thompson AM, Baxter-Jones AD, Mirwald RL, Bailey DA. Comparison of physical activity in male and female children: Does maturation matter? Med Sci Sports Exerc 2003;35:1684–90.
63. Katzmarzyk PT, Malina RM. Contribution of organized sports participation to estimated daily energy expenditure in youth. Pediatr Exerc Sci 1998;10:378–86.
64. Huang YC, Malina RM. Physical activity and health-related physical fitness in Taiwanese adolescents. J Physiol Anthropol Appl Human Sci 2002;21:11–9.
65. Andersen KL, Ilmarinen J, Rutenfranz J, Ottmann W, Berndt I, Kylian H, Ruppel M. Leisure time sport activities and maximal aerobic power during late adolescence. Eur J Appl Physiol Occup Physiol 1984;52:431–6.
66. Krahenbuhl GS, Skinner JS, Kohrt WM. Developmental aspects of maximal aerobic power in children. Exerc Sport Sci Rev 1985;13:503–38.
67. Armstrong N, Welsman JR. Assessment and interpretation of aerobic fitness in children and adolescents. Exerc Sport Sci Rev 1994;22:435–76.

68. Pate RR, Dowda M, Ross JG. Associations between physical activity and physical fitness in American children. Am J Dis Child 1990;144:1123–9.
69. Sallis JF, McKenzie TL, Alcaraz JE. Habitual physical activity and health-related physical fitness in fourth-grade children. Am J Dis Child 1993;147:890–6.
70. Schmucker B, Rigauer B, Hinrichs W, Trawinski J. Motor abilities and habitual physical activity in children. In: Ilmarinen J, Valimaki I, eds. Children and sport. Berlin: Springer Verlag, 1984:46–52.
71. Aaron DJ, Kriska AM, Dearwater SR, Anderson RL, Olsen TL, Cauley JA, Laporte RE. The epidemiology of leisure physical activity in an adolescent population. Med Sci Sports Exerc 1993;25:847–53.
72. Renson R, Beunen G, Claessens AL, Colla R, Lefevre J, Ostyn M, Schueremans C, Simons J, Taks M, Van Gerven D. Physical fitness variation among 13 to 18 year old boys and girls according to sport participation. In: Beunen G, Ghesquiere J, Reybrouck T, Claessens AL, eds. Children and exercise. Stuttgart, Germany: Ferdinand Enke Verlag, 1990:136–44.
73. Katzmarzyk PT, Malina RM, Song TM, Bouchard C. Physical activity and health-related fitness in youth: a multivariate analysis. Med Sci Sports Exerc 1998;30: 709–14.
74. Talbot LA, Metter EJ, Fleg JL. Leisure-time physical activities and their relationship to cardiorespiratory fitness in healthy men and women 18–95 years old. Med Sci Sports Exerc 2000;32:417–25.
75. Reeves L, Broeder CE, Kennedy-Honeycutt L, East C, Matney L. Relationship of fitness and gross motor skills for five- to six-yr.-old children. Percept Mot Skills 1999;89:739–47.
76. Okely AD, Booth ML, Patterson JW. Relationship of cardiorespiratory endurance to fundamental movement skill proficiency among adolescents. Pediatr Exerc Sci 2001;13:380–91.
77. Katzmarzyk PT, Malina RM, Song TMK, Bouchard C. Television viewing, physical activity, and health-related fitness of youth in the Quebec Family Study. J Adolesc Health 1998;23:318–25.
78. Gortmaker SL, Must A, Sobol AM, Peterson K, Colditz GA, Dietz WH. Television viewing as a cause of increasing obesity among children in the United States, 1986–1990. Arch Pediatr Adolesc Med 1996;150:356–62.
79. Crespo CJ, Smit E, Troiano RP, Bartlett SJ, Macera CA, Andersen RE. Television watching, energy intake, and obesity in US children: results from the third National Health and Nutrition Examination Survey, 1988–1994. Arch Pediatr Adolesc Med 2001;155:360–5.
80. Blair SN, Clark DG, Cureton KJ, Powell KE. Exercise and fitness in childhood: implications for a lifetime of health. In: Gisolfi CV, Lamb DR, eds. Perspectives in exercise science and sports medicine. II. Youth, exercise, and sport. Indianapolis, Ind, USA: Benchmark Press, 1989:401–30.
81. Mirwald RL, Bailey DA. Maximal aerobic power. London, Ont, Canada: Sport Dynamics, 1986.
82. Beunen GP, Malina RM, Renson R, Simons J, Ostyn M, Lefevre J. Physical activity and growth, maturation and performance: a longitudinal study. Med Sci Sports Exerc 1992;24:576–85.
83. Verschuur R. Daily physical activity: longitudinal changes during the teenage period. Haarlem, Netherlands: Uitgeverrij de Vrieseborch, 1987.
84. Spurr GB. Physical activity and energy expenditure in undernutrition. Prog Food Nutr Sci 1990;14:139–92.
85. Spurr GB, Reina JC. Undernutrition, physical activity, and performance of children. In: Blimkie CJR, Bar-Or O, eds. New horizons in pediatric exercise science. Champaign, Ill, USA: Human Kinetics, 1995:149–59.
86. Satyanarayana K, Naidu AN, Rao BS. Nutritional deprivation in childhood and the body size, activity, and physical work capacity of young boys. Am J Clin Nutr 1979;32:1769–75.
87. Spurr GB, Reina JC, Dahners HW, Barac-Nieto M. Marginal malnutrition in school-aged Colombian boys: functional consequences in maximum exercise. Am J Clin Nutr 1983;37:834–47.
88. Malina RM, Buschang PH. Growth, strength and motor performance of Zapotec children, Oaxaca, Mexico. Hum Biol 1985;57:163–81.
89. Benefice E, Malina RM. Body size, body composition and motor performances of mild-to-moderately undernourished Senegalese children. Ann Hum Biol 1996;23:307–21.
90. Benefice E, Fouere R, Malina RM. Early nutritional history and motor performance of Senegalese children, 4–6 years of age. Ann Hum Biol 1999;26:443–55.
91. Fellmann N, Coudert J. Malnutrition and anaerobic performance in children. In: Van Praagh E, ed. Pediatric anaerobic performance. Champaign, Ill, USA: Human Kinetics, 1998:319–35.
92. Malina RM, Little BB, Shoup RF, Buschang PH. Adaptive significance of small body size: strength and motor performance of school children in Mexico and Papua New Guinea. Am J Phys Anthropol 1987;73:489–99.
93. Malina RM, Katzmarzyk PT, Siegel SR. Overnutrition, undernutrition and the body mass index: implications for strength and motor fitness. In: Parizkova J, Hills AP, eds. Physical fitness and nutrition during growth. Basel, Switzerland: Karger, 1998:13–26.
94. Stephenson LS, Latham MC, Adams EJ, Kinoti SN, Pertet A. Physical fitness, growth and appetite of Kenyan school boys with hookworm, *Trichuris trichiura* and *Ascaris lumbricoides* infections are improved four months after a single dose of albendazole. J Nutr 1993;123:1036–46.
95. Adams EL, Stephenson LS, Latham MC, Kinoti SN. Physical activity and growth of Kenyan school children with hookworm, *Trichuris trichiura* and *Ascaris lumbricoides* infections are improved after treatment with albendazole. J Nutr 1994;124:1199–1206.
96. Stephenson LS, Latham MC, Kurz KM, Miller D, Kinoti SN, Odouri ML. Urinary iron loss and physical fitness of Kenyan children with urinary schistosomiasis. Am J Trop Med Hyg 1985;34:322–30.
97. Neumann C, McDonald MA, Sigman M, Bwibo N. Medical illness in school-age Kenyans in relation to nutrition, cognition, and playground behaviors. J Dev Behav Pediatr 1992;13:392–8.
98. Seckler D. "Malnutrition": an intellectual odyssey. West J Agricult Econ 1980;5:219–27.
99. Pelto GH, Pelto PJ. Small but healthy? An anthropological perspective. Hum Organ 1989;48:11–5.
100. Martorell R. Body size, adaptation and function. Hum

Organ 1989;48:15–20.
101. Bandini LG, Schoeller DA, Dietz WH. Energy expenditure in obese and nonobese adolescents. Pediatr Res 1990;27:198–203.
102. Shah M, Jeffery RW. Is obesity due to overeating and inactivity, or to a defective metabolic rate? A review. Ann Behav Med 1991;13:73–81.
103. Manos TM, Gutin B, Rhodes T, Spandorfer PR, Jackson LW, Litaker MS. Energy expenditure and intake in obese and nonobese African-American girls. Ann New York Acad Sci 1993;699:275–7.
104. Treuth MS, Figueroa-Colon R, Hunter GR, Weinsier RL, Butte NF, Goran MI. Energy expenditure and physical fitness in overweight vs non-overweight prepubertal girls. Int J Obes Relat Metab Disord 1998;22:440–7.
105. Goran MI. Energy expenditure, body composition, and disease risk in children and adolescents. Proc Nutr Soc 1997;56(1B):195–209.
106. Lazzer S, Boirie Y, Bitar A, Montaurier C, Vernet J, Meyer M, Vermorel M. Assessment of energy expenditure associated with physical activities in free-living obese and nonobese adolescents. Am J Clin Nutr 2003; 78:471–9.
107. Bar-Or O, Foreyt J, Bouchard C, Brownell KD, Dietz WH, Ravussin E, Salbe AD, Schwenger S, St Jeor S, Torun B. Physical activity, genetic, and nutritional considerations in childhood weight management. Med Sci Sports Exerc 1998;30:2–10.
108. Dionne I, Almeras N, Bouchard C, Tremblay A. The association between vigorous physical activities and fat deposition in male adolescents. Med Sci Sports Exerc 2000;32:392–5.
109. Trost SG, Sirard JR, Dowda M, Pfeiffer KA, Pate RR. Physical activity in overweight and nonoverweight preschool children. Int J Obes Relat Metab Disord 2003;27:834–9.
110. Malina RM, Beunen GP, Classens AL, Lefevre J, Vanden Eynde B, Renson R, Vanreusel B, Simons J. Fatness and physical fitness of girls 7 to 17 years. Obes Res 1995;3:221–31.
111. Beunen G, Malina RM, Ostyn M, Renson R, Simons J, Van Gerven D. Fatness, growth and motor fitness of Belgian boys 12 through 20 years of age. Hum Biol 1983;55:599–613.
112. Deforche B, Lefevre J, de Bourdeaudhuij I, Hills AP, Duquet W, Bouckaert J. Physical fitness and physical activity in obese and nonobese Flemish youth. Obes Res 2003;11:434–41.
113. Blimkie CJ, Ebbesen B, MacDougall D, Bar-Or O, Sale D. Voluntary and electrically evoked strength characteristics of obese and nonobese preadolescent boys. Hum Biol 1989;61:515–32.
114. Blimkie CJ, Sale DG, Bar-Or O. Voluntary strength, evoked twitch contractile properties and motor unit activation of knee extensors in obese and non-obese adolescent males. Eur J Appl Physiol Occup Physiol 1990; 61:313–8.
115. Mocellin R, Rutenfranz J. Investigations of the physical working capacity of obese children. Acta Paediatr Scand Suppl 1971;217:77–9.
116. Bar-Or O, Rowland TW. Pediatric exercise medicine: from physiologic principles to health care application. Champaign, Ill, USA: Human Kinetics, 2004.
117. Cooper DM, Poage J, Barstow TJ, Springer C. Are obese children truly unfit? Minimizing the confounding effect of body size on the exercise response. J Pediatr 1990;116:223–30.
118. DeMeersman RE, Stone S, Schaefer DC, Miller WW. Maximal work capacity in prepubescent obese and nonobese females. Clin Pediatr 1985;24:199–200.
119. Rowland TW. Effects of obesity on aerobic fitness in adolescent females. Am J Dis Child 1991;145:764–8.
120. Prentice AM, Jebb SA. Obesity in Britain: gluttony or sloth? BMJ 1995;311:437–9.
121. Risser W, Ward S. USA Today snapshots: How do Americans get to work? USA Today 2001;May 8:1A.
122. Murray JP, Kippax S. Television diffusion and social behaviour in three communities: a field experiment. Austral J Psychol 1977;29:31–43.
123. Williams TM, Handford AG. Television and other leisure activities. In: Williams TM, ed. The impact of television: a natural experiment in three communities. New York: Academic Press, 1986:143–213.
124. Ratnesar R. The homework ate my family. Time 1999;Jan 25:54–63.
125. Winters R. Your assignment in 2004. Time 1999;Jan 25:60–1.
126. Sturm R. Childhood obesity—What we can learn from existing data on societal trends, Part 1. Prev Chron Dis (serial on line) 2005;2(1):1–9. Available at www.cdc.gov/pcd/issues/2005/jan/04_0038.htm. Accessed 13 August 2006.
127. Eisenmann JC, Katzmarzyk PT, Tremblay MS. Leisure-time physical activity levels among Canadian adolescents, 1981–1998. J Phys Act Health 2004;1:154–62.
128. Lowry R, Brener N, Lee S, Epping J, Fulton J, Eaton D. Participation in high school physical education—United States, 1991–2003. MMWR 2004;53:844–7.
129. Malina RM. Secular changes in growth, maturation, and physical performance. Exerc Sport Sci Rev 1978;6: 203–55.
130. Heeboll-Nielsen K. Muscle strength of boys and girls, 1981 compared to 1956. Scand J Sports Sci 1982;4: 37–43.
131. Godina EZ. Secular changes in Russia and the former Soviet Union. In: Bodzsar EB, Susanne C, eds. Secular growth changes in Europe. Budapest, Hungary: Eotvos University Press, 1998:351–67.
132. Bliss JG. A study of progression based on age, sex, and individual differences in strength and skill. Am Phys Educ Rev 1927;32:11–21, 85–9.
133. Malina RM. Fitness and performance: adult health and the culture of youth. In: Parks RJ, Eckert HM, eds. New possibilities, new paradigms? American Academy of Physical Education Papers No. 24. Champaign, Ill, USA: Human Kinetics, 1991:30–8.
134. Mekota K. Secular trend in physical fitness in Czechoslovak school population. In: Proceedings of IV Congress of Sports Pedagogues of Yugoslovia and I International Symposium: Sport of the Young. Ljubljana-Bled, Yugoslavia: Faculty of Physical Culture, University of Ljubljana, 1990:321–3.
135. Pohndorf RH. Our unfit U.S. school children. Newsletter, Wisconsin Association for Health, Physical Education and Recreation 1961;31(May):1–4.
136. Updyke WF, Willett MS. Physical fitness trends in

American youth. Bloomington, Ind, USA: Chrysler AAU Physical Fitness Program, Indiana University, 1989.
137. Pate RR, Ross JG, Dotson CO, Gilbert GG. The new norms: a comparison with the 1980 AAHPERD norms. J Phys Educ Rec Dance 1985;56:28–30.
138. Raczek J. Entwicklungsveranderungen der motorischen Leistungsfahigkeit der Schuljugend in drei Jahrehnten (1965–1995). Sportwissenschaft 2002;32:201–16.
139. Westerstahl M, Barnekow-Bergkvist M, Hedberg G, Jansson E. Secular trends in body dimensions and physical fitness among adolescents in Sweden from 1974 to 1995. Scand J Med Sci Sports 2003;13:128–37.
140. Dollman J. Trends and sociodemographic distribution of children's health-related fitness and behaviours. Doctoral dissertation, University of South Australia, Mawson Lakes Campus, 2003.
141. Nishijima T, Kokudo S, Ohsawa S. Changes over the years in physical and motor ability in Japanese youth in 1964–97. Int J Sport Health Sci 2003;1:164–70.
142. Nishijima T, Nakano T, Takahashi S, Suzuki K, Yamada H, Kokudo S, Ohsawa S. Relationship between changes over the years in physical ability and exercise and sports activity in Japanese youth. Int J Sport Health Sci 2003;1:110–8.
143. Shingo N, Takeo M. The educational experiments of school health promotion for the youth of Japan: analysis of the 'sport test' over the past 34 years. Health Promot Int 2002;17:147–60.
144. Tomkinson GR, Olds TS, Gulbin J. Secular trends in physical performance of Australian children. Evidence from the Talent Search Program. J Sports Med Phys Fitness 2003;43:90–8.
145. Tomkinson GR, Leger LA, Olds TS, Cazorla G. Secular trends in the performance of children and adolescents (1980–2000): an analysis of 55 studies of the 20 m shuttle run test in 11 countries. Sports Med 2003;33:285–300.
146. Eisenmann JC, Malina RM. Secular change in peak oxygen consumption among United States youth in the 20th century. Am J Hum Biol 2002;14:699–706.
147. Malina RM. Cardiovascular health status of Latin American children and youth. In: Blimkie CJR, Bar-Or O, eds. New horizons in pediatric exercise science. Champaign, Ill, USA: Human Kinetics, 1995:195–220.
148. Lobstein T, Baur L, Uauy R; IASO International Obesity TaskForce. Obesity in children and young people: a crisis in public health. Obes Rev 2004;5(suppl 1):4–104.
149. Tremblay MS, Willms JD. Secular trends in the body mass index of Canadian children. CMAJ 2000;163:1429–133 (published erratum CMAJ 2001;164:970).
150. Tremblay MS, Katzmarzyk PT, Willms JD. Temporal trends in overweight and obesity in Canada, 1981–1996. Int J Obes Relat Metab Disord 2002;26:538–43.
151. Ogden CL, Flegal KM, Carroll MD, Johnson CL. Prevalence and trends in overweight among US children and adolescents, 1999–2000. JAMA 2002;288:1728–32.
152. Flegal KM, Troiano RP. Changes in the distribution of body mass index of adults and children in the US population. Int J Obes Relat Metab Disord 2000;24:807–18.
153. Hedley AA, Ogden CL, Johnson CL, Carroll MD, Curtin LR, Flegal KM. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. JAMA 2004;291:2847–50.
154. Hulens M, Beunen G, Claessens AL, Lefevre J, Thomis M, Philippaerts R, Borms J, Vrijens J, Lysens R, Vansant G. Trends in BMI among Belgian children, adolescents and adults from 1969 to 1996. Int J Obes Relat Metab Disord 2001;25:395–9.
155. Cook S, Weitzman M, Auinger P, Nguyen M, Dietz WH. Prevalence of a metabolic syndrome phenotype in adolescents: findings from the third National Health and Nutrition Examination Survey, 1988–1994. Arch Pediatr Adolesc Med 2003;157:821–7.
156. de Ferranti SD, Gauvreau K, Ludwig DS, Neufeld EJ, Newburger JW, Rifai N. Prevalence of the metabolic syndrome in American adolescents: findings from the Third National Health and Nutrition Examination Survey. Circulation 2004;110:2494–7.
157. Strong WB, Malina RM, Blimkie CJ, Daniels SR, Dishman RK, Gutin B, Hergenroeder AC, Must A, Nixon PA, Pivarnik JM, Rowland T, Trost S, Trudeau F. Evidence based physical activity for school-age youth. J Pediatr 2005;146:732–7.
158. Katzmarzyk PT, Malina RM, Bouchard C. Physical activity, physical fitness, and coronary heart disease risk factors in youth: the Quebec Family Study. Prev Med 1999;29(6 pt 1):555–62.
159. Brage S, Wedderkopp N, Ekelund U, Franks PW, Wareham NJ, Anderson LB, Froberg K; European Youth Heart Study (EYHS). Features of the metabolic syndrome are associated with objectively measured physical activity and fitness in Danish children: the European Youth Heart Study (EYHS). Diabetes Care 2004;27:2141–8.
160. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: How many days of monitoring are needed? Med Sci Sports Exerc 2000;32:426–31.
161. Sturm R. Childhood obesity—What we can learn from existing data on societal trends, Part 2. Prev Chron Dis (serial on line) 2005;2(2):1–9. Available at: www.cdc.gov/pcd/issues/2005/apr/04_0039.htm. Accessed 13 August 2006.
162. Chaing J, Wu HC, Shih DS. The 1997 nation-wide children and youth fitness study. Taipei, Taiwan: Ministry of Education, 1998 (in Chinese).

Body composition assessment for development of an international growth standard for preadolescent and adolescent children

Timothy G. Lohman and Scott B. Going

Abstract

General considerations in assessing body composition in children and youths are described. Various methods are explored and recommendations are made for methods to be used in the International Growth Standard for Preadolescent and Adolescent Children Project. Exclusion of under- and overweight participants is recommended, and a method is proposed to assess both underweight and overweight. In addition to height and weight, we recommend waist circumference, selected skinfolds, and dual-energy x-ray absorptiometry (DXA) as a measure of fat, lean, and bone mineral density. We also propose using both fat mass index and fat-free mass index as an improvement over body-mass index.

Key words: Adolescent, anthropometry, body composition, body fat, body-mass index, DXA, growth, preadolescent, total body water

What to measure?

In this chapter we consider in vivo measures of body composition for the inclusion of healthy participants and exclusion of under- and overweight participants in an International Growth Standard for Preadolescent and Adolescent Children. First we discuss general considerations in assessing body composition in children and youths. Secondly, we review various methods for assessing body composition in children and then conclude with a discussion of the issue of using various criteria to establish a healthy sample.

Timothy G. Lohman is affiliated with the Department of Physiology, and Scott B. Going is affiliated with the Department of Nutritional Sciences, University of Arizona, Tucson, Arizona, USA.

Please direct queries to the corresponding author: Timothy G. Lohman, Department of Physiology, Ina Gittings Building 93, University of Arizona, Tucson, AZ 85721, USA; e-mail: lohman@u.arizona.edu.

General considerations in assessing body composition

Age and maturation

A consistent set of measurements across preadolescent and adolescent populations is an essential component for the development of an international growth standard for preadolescent and adolescent children ranging in age from 5 to 17 years. Each method must be applicable to the entire population and must be administered using a study-wide protocol. Body-mass index, skinfold thicknesses, and circumferences vary in their relation to body fatness with age and maturation. To account for age and, to some extent, for maturation, percentiles within age have been developed as a relative measure of body composition.

Body composition

The operational definition of body composition for the purpose of the proposed international growth standard needs to be stated explicitly. Many of the simplest methods are based on the two-component model that resolves weight into fat mass (FM) and fat-free mass (FFM). There are several methods of estimating FM or FFM (the remaining component is calculated by difference from body weight). The advantage of this approach is that a single method can be used to develop a reference standard, although only whole-body fatness is estimated. If fat distribution (known to be associated with metabolic dysregulation) is to be included, then several estimates of fatness will have to be included, such as leg, trunk and arm, or abdominal fat within the truncal area or intraabdominal fat within the abdomen. The distribution of subcutaneous fat has also been shown to be related to risk factors for chronic disease and would require multiple measures at selected sites. Muscle and skeletal weight can also be estimated if a further examination of FFM is to be pursued. Since muscle mass and skeletal mass are only moderately related, independent measures of each are needed for the development of a growth standard. Skeletal weight can be further characterized by assessments of regional

bone mass and density throughout the body, based on estimates of bone mass and bone density at selected bone sites.

Sex and ethnic or racial background

At different ages and stages of maturation, differences in body composition exist between males and females and among various racial and ethnic groups. Further, the relationship between body-mass index (BMI) and body composition and between skinfolds and body composition varies across racial groups and ethnicity.

Measuring fat and fat-free body mass (two-component model)

Densitometry

Densitometry, although historically considered a criterion method, is not recommended across the age range of interest because of technical challenges and biological variability in FFM components that invalidate model assumptions. Hydrostatic weighing, the most common approach, requires accurate measures of underwater weight and lung gas volume at the time of weighing to estimate body density. The most accurate measures are obtained during complete submersion after maximal exhalation, with simultaneous measurement of functional residual capacity. This is difficult for young children, who may be uncomfortable underwater, especially if required to exhale before submerging. Partial submersion and alternative lung volumes have been tried in adults, with mixed success [1]. Although healthy, motivated adults can reasonably match lung volumes in and out of the water, it is more difficult for young children to do so. The relatively recent introduction of the Bod Pod, based on air displacement plethysmography, overcomes some of the challenges of hydrostatic weighing, since submersion in water is not required. Although somewhat more costly than the apparatus needed for hydrostatic weighing, the Bod Pod is affordable and is reasonably compact and portable, and could be mounted on a mobile laboratory. The subject chamber, with its Plexiglas window, is comfortable for most children. To measure thoracic volume, a "puffing" maneuver against a closed valve is required, which may be difficult for younger children to perform. Good to excellent reliability has been reported, although the results in children are not always as good as in adults [1–3]. The average agreement between Bod Pod and underwater weighing or dual-energy x-ray absorptiometry (DXA) is generally good (approximately 1% to 2% body fat), although wide intersubject variations occur [2] that are perhaps related to body size, age, or other factors [3].

In addition to the measurement errors associated with obtaining the estimate of density, additional error is accrued, which is related to the validity of the assumptions underlying the equations used to estimate percent fat and FFM from body density, regardless of the method used to estimate density. The changes in the FFM fractions of protein, water, and mineral that occur with growth and maturation and their impact on estimates of percent fat and FFM have been well described. Application of adult equations in children can lead to errors in the estimation of percent fat on the order of 4% to 8% or more, depending on maturation [4]. The application of population-specific equations, generally based on age-specific estimates of average FFM composition, significantly reduces the average error, although errors for individuals at the extremes of maturation at a given age remain substantial. Multicomponent models, whereby density is adjusted for the water and mineral fractions of FFM, provide more accurate estimates of composition but add additional cost and technical complexity by requiring estimates of additional components of the FFM [1].

Body water

Body water was first recommended for estimation of body composition by Pace and Rathburn [5] on the basis of chemical analyses of several animal species. Schoeller [6] recently reviewed methods for measuring body water using various tracer approaches, describing the advantages and limitations of each approach, and proposed a standardized protocol to maximize the accuracy and precision of body water estimates from the isotope dilution approach. Total body water (TBW) can be estimated from saliva, urine, or blood samples, and this dilution approach is practical in large samples of subjects. The use of stable isotopes (deuterium and oxygen-18) makes this method applicable to children of all ages and allows estimation of FFM and FM.

The use of body water to estimate FFM assumes a constant water content of the FFM among healthy children of various ages. In a review of this constant by Wang et al. [7], substantial evidence supports an average hydration between 72% and 74% in a variety of species, including adult humans. There is considerable evidence that prepubescent and pubescent children between 6 and 14 years of age have a higher hydration level than young adults. The water content of FFM decreases with age and maturation at the same time as the bone mineral content of FFM increases, with chemical maturity reached in the postpubescent years. Early work by Fomon et al. [8], Haschke et al. [9], and Haschke [10] developed the concept of changes in reference body composition for children and adolescents. Further work with direct measures of water and bone mineral by Boileau et al. [11], Lohman et al. [12], and Slaughter et al. [13] confirmed the chemical immaturity of prepubescent children. Boileau et al. [14] and Lohman [15] summarized this work and that of others to further establish formulas and constants for body composition estimates in children. Lohman [16] extended this research to the densities of FFM

in children from infants to adolescents and defined age-specific formulas for use in children [16], rather than the often-used Siri equation intended for adults. Deurenberg et al. [17] also worked with the data of Fomon et al. [8] to develop age-adjusted equations for body density and found estimates of body composition in children comparable to those made by Lohman [18] and Slaughter et al. [19]. Although the impact of chemical immaturity in prepubescent children and the potential inaccuracies of two-component models in children based on underwater weighing and air displacement plethysmography are generally recognized [2, 20–24], the use of multicomponent models in body composition studies of children is not universal [25–27]. Of the available two-component approaches, estimation of FFM from TBW is preferred, as long as the appropriate (age- or maturation-adjusted) conversion constants are used. Isotope and sample processing as well as analytical procedures are potential challenges for standardization among countries.

Dual-energy x-ray absorptiometry (DXA)

DXA provides an ideal method to assess body composition in children and youths [15]. Radiation exposure is low, and unlike other methods that are confounded by changes in FFM composition during growth and development (e.g., body density, body water, bioelectrical impedance [BIA], and anthropometry), DXA estimates of bone mineral, soft tissue, and fat are not greatly affected by the hydration level of FFM. Theoretical analyses in adults [28] and in children from infancy to 10 years of age [29] have shown that the typical variation in FFM hydration affects DXA estimates of percent fat by less than 1% during growth and development. The disadvantage of DXA is the variation in hardware and software over the past 10 years, rendering much of the past research out of date.

Ellis et al. [30] have developed reference models for children and adolescents using DXA in a population of black, white, and Hispanic children from 5 to 19 years of age. In general, their results confirm older estimates of FFM and FM by Fomon et al. [8] and Haschke [31] using indirect estimates of body composition from the literature. Sopher et al. [32] recently compared DXA with a four-component model adjusted for mineral and water in a large sample of children and adolescents ($N = 411$) 6 to 18 years of age. Close average (approximately 1%) agreement was found between methods; they were highly correlated ($R^2 = 0.85$), and prediction accuracy was good (SEE = 3.7%). In this sample, DXA overestimated percent fat in children with more than 30% fat by 3% to 5% and underestimated percent fat in children with less than 10% fat by 1% to 2%. Because of the improvement in both hardware and software for DXA over the past several years, we recommend this method with the use of the Lunar Prodigy fan-beam approach in all samples.

Skinfolds and subcutaneous fat patterns

The use of skinfolds to estimate percent fat and, indirectly, FFM in children and adults is well established. Many studies as well as national nutrition surveys have included skinfolds along with height and weight to better describe changes in body fat with growth and development. Leading candidates for skinfold sites include the triceps and subscapular sites as representative measures of extremity and trunk subcutaneous fat depots. Skinfold norms for many countries have been developed, defining the 50th, 85th, and 95th percentiles. Secular trends over the past 20 years have demonstrated increases in both median skinfold thicknesses and the number of children with higher values.

Studies relating skinfolds to whole-body fatness in children using two-, three-, and four-component models have shown a prediction error of 3% to 4% body fat (standard error of estimate), and many equations have been developed to assess body fat from skinfold measurements [15, 17, 19, 33, 34]. Roemmich et al. [21] emphasized the need for multicomponent models as criterion methods for the development of skinfold equations and found that equations based on three- or four-component models were more valid than those based on two-component models, which lead to a large systematic overestimation (≥ 5%) of body fat. Additional skinfold formulas have recently been developed using DXA as the criterion method [35].

Given that the relationship between skinfolds and body fat varies with age, maturation [19, 36], obesity, and ethnicity [37], there have been few attempts to develop universal body fat norms derived from skinfolds for children [4]. Using the National Center for Health Statistics (NCHS) skinfold data for skinfolds [38], Lohman [4] derived percent fat norms (**table 1**) based on converting skinfolds to percent fat using the Slaughter et al. formula [19]. He showed an increase in skinfold thicknesses from the National Health Examination Survey (NHES) [38] in 1973 to the National Children and Youth Fitness Study (NCYFS) [39] in 1985 in children 6 to 16 years of age [4]. Because fat pattern also changes with age and maturation, so that truncal fat and abdominal fat increase with age, and given the association between abdominal fatness and disease risk factors [40], it is important to estimate regional as well as total body fatness, especially in late adolescence [41–43]. Subscapular and abdominal skinfolds are good candidates for measurement of truncal fat and need to be included if changes in fat patterning are to be assessed in a survey. A standardized set of skinfolds and circumferences has been developed, and research over the past 10 years [44] has used a more uniform set of procedures to assess body composition, making skinfolds an important candidate for both body fat estimates and estimates of fat patterning with growth. We recommend measuring triceps, subscapular, and abdominal skinfolds for the

TABLE 1. Percent fat norms derived from National Health Examination Survey (NHES) skinfold data[a]

Age (yr)	Males[b]						Females[c]					
	50th percentile		85th percentile		95th percentile		50th percentile		85th percentile		95th percentile	
	Σ2SK	% fat	Σ2SK	% fat	Σ2SK	% fat	Σ2SK	% fat	Σ2SK	% fat	Σ2SK	% fat
6	12	11.7	16	15.6	20	19.6	14	13.5	19	18.1	27	23.9
8	13	12.7	19	18.4	28	25.9	16	15.5	25	22.6	36	29.4
10	14	12.7	24	21.9	33	28.8	18	17.2	31.6	26.5	43	33.2
12	14.5	12.5	27	23.6	44	35.0	19.5	18.5	34	27.7	47.3	35.5
14	14	11.0	26.5	22.0	39	30.5	23.5	21.6	37.5	30.2	52.6	38.4
16	14	9.0	24	18.9	39	29.3	25.5	23.0	42	32.6	58	41.4

a. Norms were derived from National Health Examination Survey (NHES) data for 1963-65 [38] using the skinfold equations from Slaughter et al. [19].

b. For males, intercept (I) varies with age. For 6- and 8-year-olds, I = −1.7; for 10-year-olds, I = −2.5; for 12-year-olds, I = −3.4; for 14-year-olds, I = −4.4; and for 16-year-olds, I = −5.5. The equation is % fat = 1.21 Σ2SK−0.008 (Σ2SK) 2 + I. For males with skinfolds greater than 35 mm, % fat = 0.783 (Σ2SK) + 2.2 (6- to 8-year-olds), 0.6 (10- to 12-year-olds), or −1.2 (14- to 16-year-olds).

c. For females, % fat = 1.33 (Σ2SK)−0.013 (Σ2SK) 2−2.5 (one intercept for all ages). For females with skinfolds greater than 35 mm, % fat = 0.546 (Σ2SK) + 9.7, for all ages.

International Growth Standard for Preadolescent and Adolescent Children.

Circumferences

Because intersubject variation in circumferences reflects variation in muscle, fat, and bone, circumferences have limited utility when applied across a broad age range including different levels of maturation. Leg circumference and especially upper-arm circumference, when corrected for individual differences in subcutaneous fat, are useful for estimating muscle plus bone cross-sectional area. Since there is greater variation in muscle than in bone, corrected upper-arm area is a relatively good index of muscle and has been often used as an index of nutritional status and growth. The approach has merit, given the importance of describing changes in different components of body composition during growth and development. Muscle is the largest tissue component of FFM, and variation in muscle due to age, maturation, sex, race or ethnicity, nutrition, and activity certainly contributes to variation in BMI and confounds its interpretation as an index of fatness across different groups.

The waist circumference may be particularly useful, given the interest in abdominal fatness and its association with metabolic dysregulation and chronic disease risk [40]. The association is much better established in adults, although emerging evidence suggests abdominal fatness is also a risk factor in children and adolescents. Use of waist circumference as an index of fatness avoids some of the problems associated with interpretation of BMI, since variation in waist circumference is expected to reasonably track variation in fatness. Moreover, given that the abdominal fat depot (especially visceral fat) may be particularly labile, assessment of waist circumference may prove beneficial for tracking changes in fatness due to intervention as well as normal growth and development. Only limited normative data exist for children and adolescents [45, 46], and certainly more work is needed to better define the relationship between waist circumference, intraabdominal and subcutaneous fatness, and chronic disease risk factors in children and adolescents. Nevertheless, given the emergence of waist circumference as an important anthropometric indicator of disease risk in adults, its inclusion in a growth study merits serious consideration.

Bioelectrical impedance analysis

Bioelectrical impedance analysis (BIA) is a simple, relatively inexpensive, and easily portable method for estimating TBW, FFM, and percent fat in field, clinical, and laboratory settings. Numerous studies show that the basic measure, bioelectrical resistance (r), and thus the derived resistance index (ht^2/r), are very reproducible. The resistance index is highly correlated with TBW and therefore with FFM and gives valid estimates of both components. The average agreement between BIA-derived estimates of TBW and FFM and criterion measures is generally good, and a substantial part of the disagreement can often be attributed to errors in the criterion method. However, even in studies with criterion methods (e.g., DXA) that are relatively unaffected by model error, errors for the individual can be substantial, and this is a limitation of the method for screening when individual values matter more than the population average. The typical single-frequency tetrapolar electrode arrangement approach assumes normal hydration and is best applied after an 8- to 12-hour fast. Other factors, such as exercise, temperature, and regional blood flow, which are difficult to control in the field, can contribute to technical error. The method is also based on "geometric" assumptions (e.g., the trunk, arms, and legs are cylinders) that may not be equally valid across populations. Although numerous validation studies have been conducted in children and adolescents [4, 37, 47], there have been far fewer cross-validation studies, and there is no consensus regarding which equations to use in this population. There have

been very few studies comparing different racial and ethnic groups, and the effect of race and ethnicity on BIA estimates of body composition in children and adolescents remains uncertain. In adults, differences between black and white men and women have been clearly demonstrated, and population-specific equations are recommended [48, 49]. The available data in children and adolescents suggest that racial differences develop early [1, 50]. The reason for the significant effect of race on BIA-derived estimates of composition is not clear. Racial differences in FFM composition are known to exist that contribute to model error, depending on the criterion method that is employed. However, even in studies using multiple-component-based criterion methods, racial differences persist [51]. There are also differences in body proportions (e.g., trunk and extremity lengths, skeletal widths, and appendicular muscle) that could have a significant effect on estimation of FFM and percent fat by BIA, since total body resistance is largely determined by segmental resistances in the extremities [52–57]. Whether racial differences in electrical properties exist that contribute to differences in BIA estimates of body composition is uncertain, although the work of Schoeller and Luke [48] in adults suggests it is the composition of excess weight rather than electrical properties that explains the racial effect.

Chumlea et al. [58], using the equations of Sun et al. [59] with data from the National Health and Nutrition Examination Survey (NHANES) III for weight, stature, and BIA, estimated TBW, FFM, FM, and percent body fat for 12- to 19.9-year-old non-Hispanic black males and females, non-Hispanic white males and females, and Mexican-American males and females (**tables 2** and **3**). Estimates of FFM are given in **table 2**, based on an assumption of no effect of race or ethnicity on the relationship of impedance to FFM. In recent work by Going et al. [51], black female adolescents had 1.7 kg more FFM for a given resistance index and a 3.3% difference in percent body fat compared with white and Hispanic female adolescents. The values in **tables 2** and **3** are adjusted for this systematic effect. Racial differences (black versus white) in the BIA to FFM relationship have also been reported by Schoeller and Luke [48] and Morrison et al. [50].

Body-mass index

The widespread use of BMI in nutrition and growth surveys over the last 30 years makes it an important candidate for a growth survey. Height and weight can be measured precisely and accurately in large samples with inexpensive equipment. Both measurements are noninvasive and can easily be performed in children. Height is essential to document linear growth with age and thus is an important part of any survey designed to obtain normative data. Weight gives additional information over and above height as an estimate of growth in mass and thus reflects the sum of muscle,

TABLE 2. Fat-free mass (kg) for children aged 12 to 18 years estimated from bioelectrical impedance using data from National Health and Nutrition Examination Survey (NHANES) III [58]

		Non-Hispanic white		Non-Hispanic black		Mexican-American	
Sex	Age (yr)	$\bar{X}$	SE	$\bar{X}^a$	SE	$\bar{X}$	SE
Male	12	41.8	1.0	40.9	0.9	40.3	0.9
	14	54.3	1.2	52.3	0.9	49.8	1.2
	16	57.8	1.0	55.3	0.9	53.0	0.9
Female	12	38.1	0.7	(41.0)	0.5	37.3	0.6
	14	40.4	0.6	(43.5)	0.9	37.8	0.6
	16	41.6	0.7	(43.7)	0.6	40.5	0.7

a. Numbers in parentheses are corrected for systematic effect of race (1.7 kg for African-American) on the relationship of bioelectrical impedance to fat-free mass).

TABLE 3. Percent fat for children aged 12 to 18 years estimated from bioelectrical impedance using data from National Health and Nutrition Examination Survey (NHANES) III [58]

		Non-Hispanic white		Non-Hispanic black		Mexican-American	
Sex	Age (yr)	$\bar{X}$	SE	$\bar{X}^a$	SE	$\bar{X}$	SE
Male	12	18.4	1.0	19.5	1.1	22.0	1.0
	14	18.4	1.2	17.8	0.9	18.8	1.1
	16	17.7	0.9	18.6	0.8	21.3	0.7
Female	12	24.8	1.2	(23.6)	0.8	28.6	0.8
	14	29.1	0.8	(27.6)	0.9	31.8	0.7
	16	30.7	0.9	(29.3)	0.9	33.3	0.8

a. Numbers in parentheses are corrected for 3.3% difference between African-Americans and non–African-Americans.

bone, organs, and fat. Although weight in relation to height (BMI) can be used as a measure of under- and overnutrition, it is especially valuable as a measure of undernutrition, because at low levels of BMI there is less variation in muscle, bone, and fat, since all three compartments have been depleted to arrive at the low value. Thus, we recommend the use of BMI as an index of underweight for screening out children who are malnourished.

In contrast, high levels of BMI can be reached as a result of varying amounts of muscle, bone, and fat, and thus one cannot discern the composition of the increased mass relative to height. Thus, although BMI is correlated with fatness, it has limited ability to detect excess adiposity because of its failure to detect obesity (lack of sensitivity) in 20% to 50% of children who are obese when measured by more direct methods [60–66], depending on the BMI cutoff points used to define obesity. In support of this point, Siervogel et al. [60] have documented the change in FFM and FM in relation to height as a function of maturation and leg and trunk length in the data collected in the Fels Longitudinal Study. They clearly show that variation in the relation of FFM to height increases BMI with age independently of adiposity. Thus, any survey of growth and development needs to include additional measurements along with height and weight to estimate both FM and FFM as a function of age and derive FM and FFM indices to monitor their contributions to overall BMI.

Racial and ethnic differences in the composition of weight per unit height also confound interpretation of the BMI. Several studies have shown significant racial or ethnic differences in body composition, especially in bone and muscle mass. For example, Cohn et al. [61] found that total body potassium and calcium were 5% to 10% higher in black women than in white women, indicating greater muscle mass and bone mass in blacks per unit of height. Similar findings have been reported in children and adolescents. For example, Novotny et al. [62] recently reported differences in body composition between Asian girls and other groups that persisted after adjustment for differences in maturation, and Going et al. [63] showed that these differences translated into differences in the relationship between BMI and percent body fat. Thus, population-specific percentiles of BMI for children of different ages and sexes are likely to be needed for screening in a study including different racial and ethnic groups.

Fat and FFM indices

The use of FFM/ht^2 (FFMI, fat-free-mass index) and FM/ht^2 (FMI, fat-mass index) are derived in a fashion analogous to BMI, has been proposed as an alternative to BMI when estimates of FFM or FM are available. Hattori et al. [64] first proposed these indices to adjust FM and FFM for size (height) as a better method for evaluation of growth in children. Work by Wells [65] and Maynard et al. [66] further supports the validity of this approach. Siervogel et al. [60] have clearly shown that changes in FFM/ht confounded interpretation of BMI in adolescents, especially around the time of peak height velocity (PHV). Thus, we recommend some method to separate FM from FFM, in addition to BMI, so that FM index and FFMI can be used as part of the body composition assessment [67, 68].

Bone density

Bone density (g/cm^2) can be assessed by DXA in growing children. The recent work of Horlick describes bone growth in children from 6 to 18 years of age. Prediction equations are given to estimate bone density from age, weight, and height [69]. We recommend that bone mineral density values be assessed by DXA in children of all ages.

Multicomponent model

The use of three- or four-component models to assess body composition in children represents the optimal approach, since the need for assumptions regarding FFM chemical compositions, and thus model error, are minimized. A summary of multicomponent models was recently published by Wang et al. [70]. The model proposed by Lohman and Going [71] has been widely used in research with children:

$$\%\ \text{fat} = (2.749/D_b - 0.714\ \text{TBW/BW} + 1.146\ \text{M/BW} - 2.0503) \times 100$$

where M is total body mineral, BW is body weight, TBW is total body water, and D_b is body density. If the multicomponent model calls for bone mineral rather than body mineral (Mo), such as the equation recommended by Withers et al. [72], then DXA bone mineral content can be used without conversion to body mineral:

$$\%\ \text{fat} = (-0.739\ \text{TBW/BW} + 0.947\ \text{Mo/BW} - 1.790) \times 100$$

Use of these models is generally accepted as the most valid approach for assessing body composition, because these models involve fewer assumptions since body water, body density, and bone mass (from DXA) are measured directly**.** However, they require additional expense and technical expertise, since more components must be measured, and they are susceptible to increased technical error that can overwhelm the improvement in model error. In practice, the expense and added burden of a four-component model may not be practical for the present purpose. Thus, we recommend DXA (based on a three-component model) instead of the more burdensome

four-component model, which provides simultaneous estimates of FM and FFM as well as bone mineral mass and bone density.

Screening and selection of a healthy sample

The definition of a healthy sample for an international growth standard can be based on percent body fat to establish cutoff points for both underweight and obesity. At any age, values between 25% and 35% for boys and between 30% and 40% for girls represent ranges for exclusion at the upper range of fatness. An exact point above 25% for boys and above 30% for girls that would indicate an unhealthy condition and thus should be an exclusion criterion is difficult to define. We suggest 30% for boys and 35% for girls, since these levels have been associated with elevated risk factors for cardiovascular disease in boys and girls aged 6 to 18 years [73].

Similarly, for the underweight sample (which we define as underfat and likely to be malnourished), body fat at a given age can be used to establish a cutoff point for exclusion based on excessively low body fat. There is general agreement that body fat below 5% to 10% for boys and 12% to 18% for girls is unhealthy. We recommend using height (third percentile based on Centers for Disease Control and Prevention [CDC] norms for age and sex) in combination with percent fat (below 7% for boys and 15% for girls) to define the cutoff point for underweight.

Early standards for body fatness in children and youths were based on skinfold thicknesses and percentiles from the NHANES national probability samples. With the development of skinfold equations based on multicomponent models [15, 19] and age-adjusted equations for estimating body density [44, 73, 74], valid estimates of percent body fat in children could be derived from skinfold measures. Using this approach, we compared the percent body fat at different skinfold percentiles for boys and girls of various ages (**table 1** and **fig. 1**). On the basis of these results, it is clear that the percent fat value corresponding to the 85th percentile of skinfold thickness for 6- to 8-year-olds (**fig. 1**) is considerably below the cutoff point of 25% for boys and 32% for girls that has been shown to be related to disease risk factors [73]. Thus, percentile of skinfold thickness is not an appropriate measure of obesity until a child is at least 10 years old.

More recently, health-related fatness standards have been developed that are derived from the relationship between body fatness and cardiovascular disease risk factors in children and adolescents [73]. In the long-running Bogalusa Heart Study, Berenson and colleagues [75] have shown that large skinfold thicknesses are related to higher blood lipids, lipoproteins, and blood pressure and impaired glucose tolerance in

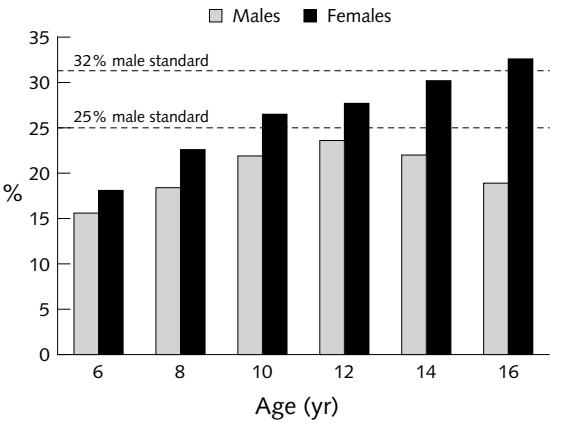

FIG. 1. Percent fat corresponding to the 85th percentile of skinfold thickness based on National Center for Health Statistics (NCHS) data [38]

children and adolescents. This study has provided a wealth of data that relate body composition in youths to the early development of cardiovascular risk factors [75–79]. Using the blood lipids and skinfold data from a subsample of 1,667 males and 1,653 females aged 5 to 18 years, Williams et al. [73] used age-adjusted, sex- and race-specific equations to estimate percent body fat from skinfold measurements and to assess the relationship between percent body fat and risk factors for cardiovascular disease. In this analysis, children were grouped by body fat (5% intervals) and odds ratios were calculated for each group (against the lowest-fat group) that reflected the risk (odds) of elevated total cholesterol, unfavorable serum lipoprotein ratios, and elevated systolic and diastolic blood pressure (**fig. 2**) [73]. Boys with greater than 25% body fat and girls with greater than 30% body fat had significantly increased risk for higher levels of adverse lipoprotein ratios and elevated blood pressure compared with their leaner counterparts (**fig. 2**). Sardinha et al. [80] have summarized cutoff points for percent fat from several other studies in the literature spanning the period from 1992 to 2001 and have shown general agreement with the findings of Williams et al. [73].

The results of Williams et al. [73] became the basis for standards setting the healthy range of body fatness at 17% to 32% for girls and 10% to 25% for boys (**figs. 3** and **4**). **Table 4** shows BMI and skinfold values corresponding with these standards that can be used to identify children above or below the healthy body fatness range. These values were established for skinfolds using the equations of Slaughter et al. [19] to give the equivalent skinfold values (triceps plus subscapular, or triceps alone). To find the equivalent BMI, the percentiles for the sum of triceps and subscapular skinfolds and for BMI were calculated from NHANES II data. The BMI at the same percentile as the sum of two skinfolds is shown in **table 4** as the cutoff point for boys (25% fat) and for girls (32% fat).

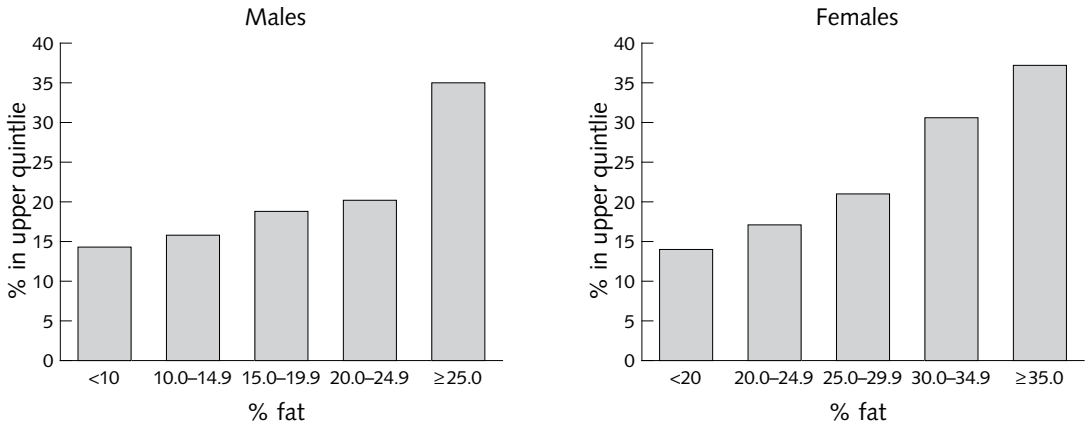

FIG. 2. Association of body fatness with cardiovascular risk factors (blood lipid and blood pressure from the results of Williams et al. [73] and Lohman [4]

Girls

% fat (all ages)	40	35	32	30	25	20	17	15	13	10	7
	Very high			High	Mod. high	Optimal range		Low		Very low	
	Needs improvement			Healthy fitness zone				Lean message			
BMI (age)											
5				21		16.2					
6				21		16.2					
7				22		16.2					
8				22		16.2					
9				23		16.2					
10				23.5		16.6					
11				24		16.9					
12				24.5		17.5					
13				24.5		17.5		14.9			
14				25		17.5		14.9			
15				25		17.5		14.9			
16				25		17.5		14.9			
17				26		17.5		14.9			
18–25				27.3		18.0		15.0			

FIG. 3. Fitnessgram health-related body composition standards for girls based on results of Williams et al. [73], with BMI cutoff points equivalent to 32% fat and 17% fat

Boys

% fat (all ages)	40	35	30	25	20	15	10
	Very high		High	Mod. high	Optimal range		Low Very low
	Needs improvement			Healthy fitness zone			Lean message
BMI (age)							
5				20	14.7		
6				20	14.7		
7				20	14.9		
8				20	15.1		
9				20	15.2		
10				21	15.3		
11				21	15.8		
12				22	16.0		
13				23	16.6	15.0	
14				24.5	17.5	15.7	
15				25	18.1	16.4	
16				26.5	18.5	16.6	
17				27	18.8	16.8	
18–25				27.8	19.0	17.0	

FIG. 4. Fitnessgram health-related body composition standards for boys based on results of Williams et al. [73], with BMI cutoff points equivalent to 25% fat and 10% fat

TABLE 4. Percent fat standards for children and adolescents with corresponding cutoff points for body-mass index (BMI) and skinfolds

Age (yr)	% Fat	Triceps and calf (subscapular) skinfolds (mm)[a]	Triceps skinfold (mm)	BMI
Males				
10	10–25	12–33 (30)	7–19	15.3–21.0
11	10–25	12–33 (30)	7–19	15.8–21.0
12	10–25	12–33 (30)	7–19	16.0–22.0
13	10–25	12–33 (30)	7–18	16.6–23.0
14	10–25	12–33 (30)	7–18	17.5–24.5
15	10–25	12–33 (30)	7–17	18.1–25.0
16	10–25	12–33 (30)	7–17	18.5–26.5
17	10–25	12–33 (30)	7–16	18.8–27.0
Females				
10	17–32	20–44 (38)	10–24	16.6–23.5
11	17–32	20–44 (38)	10–24	16.9–25.0
12	17–32	20–44 (38)	10–24	16.9–24.5
13	17–32	20–44 (38)	10–23	17.5–24.5
14	17–32	20–44 (38)	10–23	17.5–25.0
15	17–32	20–44 (38)	10–23	17.5–25.0
16	17–32	20–44 (38)	10–22	17.5–25.0
17	17–32	20–44 (38)	10–22	17.5–26.0

a. Values in parentheses are the sum of triceps and subscapular skinfolds.

A potential limitation of this approach is the dependence of percent fat on FFM, the requirement of an accurate method for estimating fat or FFM, and uncertainty in cutoff points for different racial or ethnic groups. Four-compartment models are ideal for estimating composition in very heterogeneous samples but may not be practical for the present purpose. Conversion factors representing the average hydration of FFM are reasonably well established, and estimation of FFM from TBW across groups is probably accurate, if application of dilution methods proves feasible. DXA is based on a three-component model that adjusts for variation in the mineral fraction of FFM and is relatively unaffected by the typical variation in body water. DXA also has the important advantage of giving regional as well as whole-body estimates of composition. The FMI (fat, kg/ht^2) may avoid some of the limitations of percent fat and deserves to be explored. However, we know of no studies linking FMI with disease risk factors in children and adolescents, and therefore we cannot recommend cutoff points for screening at this time. In contrast, in the analysis by Williams et al. [73], associations between composition and risk factors were similar (within sexes) in black and white boys and girls across a broad age range (6 to 18 years) when composition was defined as percent fat.

Recommendations for screening out obese children

We have shown that boys with percent body fat above 25% and girls with percent body fat above 32% are more likely to have elevated (relative to their peers) cardiovascular risk factors. We recommend that percent body fat values above 30% for boys and 35% for girls be used as the cutoff points for obesity. Boys with an estimated triceps skinfold thickness of 22 mm and girls with an estimated thickness of 27 mm would be screened out of the survey. BMI cutoff points for age corresponding to 30% fat (boys) and 35% fat (girls) can also be used, although the triceps skinfolds approach is preferred. The BMI will be slightly higher than those used in the Fitnessgram standards (**table 4**) and those proposed by various norms if 30% and 35% fat are the final cutoff points selected for boys and girls, respectively.

Recommendations for screening out underweight children

Girls with less than 15% fat and boys with less than 7% fat are at greater risk for being underweight. We recommend that all children whose height is below the third percentile (CDC norms for age and sex) and with estimated body fat below 15% for girls and 7% for boys be excluded from the survey. For girls, 15% fat corresponds to a triceps skinfold of 9 mm, and for boys 7% fat corresponds to a triceps skinfold of 5 mm. BMI cutoff points for age and sex corresponding to 15% and 7% fat can also be used to screen out underweight children, although triceps skinfold measurements are preferred. There are no published studies comparing the sensitivity and specificity of these proposed cutoff points with those of more direct indicators of underweight.

References

1. Going SB. Hydrodensitometry and air displacement plethysmography. In: Heymsfield SB, Lohman TG, Wang ZM, Going SB (eds), Human Body Composition, Second Edition. Champaign, Ill, USA: Human Kinetics, 17–33, 2005.
2. Fields DA, Goran MI, McCrory MA. Body-composition assessment via air-displacement plethysmography in adults and children: a review. Am J Clin Nutr 2002;75:453–67.
3. Demerath EW, Guo SS, Chumlea WC, Towne B, Roche AF, Siervogel RM. Comparison of percent body fat estimates using air displacement plethysmography and hydrodensitometry in adults and children. Int J Obes Relat Metab Disord 2002;26:389–97.
4. Lohman TG. Advances in body composition assessment. Champaign, Ill, USA: Human Kinetics Publishers, 1992.
5. Pace N, Rathburn EN. Studies of body composition. III. The body water and chemically combined nitrogen content in relation to fat content. J Biol Chem 1945; 158:685–91.
6. Schoeller DA. Hydrometry. In: Heymsfield SB, Lohman TG, Wang ZM, Going SB (eds). Human Body Composition, 2nd ed. Champaign, Ill, USA: Human Kinetics, 35–49, 2005.
7. Wang ZM, Deurenberg P, Wang W, Pietrobelli A, Baumgartner RN, Heymsfield SB. Hydration of fat-free body mass: review and critique of a classic body-composition constant. Am J Clin Nutr 1999;69:833–41.
8. Fomon SJ, Haschke F, Ziegler EE, Nelson SE. Body composition of reference children from birth to age 10 years. Am J Clin Nutr 1982;35:1169–75.
9. Haschke F, Fomon SJ, Ziegler EE. Body composition of a nine-year-old reference boy. Pediatr Res 1981;15: 847–9.
10. Haschke F. Body composition of adolescent males. Acta Paediatr Scand 1983;307:(suppl)1–20.
11. Boileau RA, Lohman TG, Slaughter MH, Bail TE, Going SB, Hendrix MK. Hydration of the fat free body in children during maturation. Hum Biol 1984;56:651–66.
12. Lohman TG, Slaughter MH, Boileau RA, Bunt JC, Lussier L. Bone mineral content measurements and their relation to body density in children, youth, and adults. Hum Biol 1984;56:667–79.
13. Slaughter MH, Lohman TG, Boileau RA, Stillman RJ, Van Loan M, Horswill CA, Wilmore JH. Influence of maturation on relationship of skinfolds to body density: A cross-sectional study. Hum Biol 1984;56:681–9.
14. Boileau RA, Lohman TG, Slaughter MH. Exercise and body composition in children and youth. Scand J Sports Sci 1985;7:17–27.
15. Lohman TG. Applicability of body composition techniques and constants for children and youths. Exerc Sport Sci Rev 1986;14:325–57.
16. Lohman TG. Assessment of body composition in children. Pediatr Exer Sci 1989;1:19–30.
17. Deurenberg P, Pieters JJ, Hautvast JG. The assessment of the body fat percentage by skinfold thickness measurements in childhood and young adolescence. Br J Nutr 1990;63:293–303.
18. Lohman TG. Applicability of body composition techniques and constants for children and youths. In: Holloszy JO (eds). Exercise and Sports Science Reviews. Baltimore, Md, USA: Williams & Wilkins, 325–57, 1986.
19. Slaughter MH, Lohman TG, Boileau RA, Horswill CA, Stillman RJ, Van Loan MD, Bemben DA. Skinfold equations for estimation of body fatness in children and youth. Hum Biol 1988;60:709–23.
20. Goran MI, Driscoll P, Johnson R, Nagy TR, Hunter G. Cross calibration of body composition techniques against dual energy X-ray absorptiometry in young children. Am J Clin Nutr 1996;63:299–305.
21. Roemmich JN, Clark PA, Weltman A, Rogol AD. Alterations in growth and body composition during puberty. Comparing multicompartment body composition models. J Appl Physiol 1997;83:927–35.
22. Wells JCK, Fuller NJ, Dewit O, Fewtrell MS, Elia M, Cole TJ. Four-component model of body composition in children: density and hydration of fat-free mass and comparison with simpler models. Am J Clin Nutr 1999;69:904–12.
23. Fields DA, Goran MI. Body composition techniques and the four-compartment model in children. J Appl Physiol 2000;89:613–20.
24. Gately PJ, Radley D, Cooke CB, Carroll S, Oldroyd B, Truscott JG, Coward WA, Wright A. Comparison of body composition methods in overweight and obese children. J Appl Physiol 2003;95:2039–46.
25. Nunez C, Kovera AJ, Pietrobelli A, Heshka S, Horlick M, Kehayias JJ, Wang Z, Heymsfield SB. Body composition in children and adults by air displacement plethysmography. Eur J Clin Nutr 1999;53:382–7.
26. Lockner DW, Heyward VH, Baumgartner RN, Jenkins KA. Comparison of air-displacement plethysmography, hydrodensitometry, and dual X-ray absorptiometry for assessing body composition of children 10 to 18 years of age. Ann NY Acad Sci 2000;904:72–8.
27. Phillips SM, Bandini LG, Compton DV, Naumova EN, Must A. A longitudinal comparison of body composition by total body water and bioelectrical impedance in adolescent girls. J Nutr 2003;133:1419–25.
28. Pietrobelli A, Formica C, Wang Z, Heymsfield SB. Dual-energy x-ray absorptiometry body composition model: review of physical concepts. Am J Physiol 1996;271: E941–E951.
29. Testolin CG, Gore R, Rivkin T, Horlick M, Arbo J, Wang Z, Chiumello G, Heymsfield SB. Dual-energy X-ray absorptiometry: analysis of pediatric fat estimate errors due to tissue hydration effects. J Appl Physiol 2000;89:2365–72.
30. Ellis KJ, Shypailo RJ, Abrams SA, Wong WW. The reference child and adolescent models of body composition. A contemporary comparison. Ann NY Acad Sci 2000;904:374–82.
31. Haschke F. Body composition during adolescence. In: 98th Ross Conference on Pediatric research. Body composition measurements in infants and children. Columbus, Ohio, USA: Ross Laboratories, 1989.
32. Sopher AB, Thornton JC, Wang J, Pierson RN, Jr., Heymsfield SB, Horlick M. Measurement of percentage

of body fat in 411 children and adolescents: a comparison of dual-energy X-ray absorptiometry with a four-compartment model. Pediatr 2004;113:1285–90.
33. Rolland-Cachera MF, Sempe M, Guilloud-Bataille M, Patois E, Pequignot-Guggenbuhl F, Fautrad V. Adiposity indices in children. Am J Clin Nutr 1982;36:178–84.
34. Boileau RA, Wilmore JH, Lohman TG, Slaughter MH, Riner WF. Estimation of body density from skinfold thicknesses, body circumferences and skeletal widths in boys aged 8 to 11 years: comparison of two samples. Hum Biol 1981;53:575–92.
35. Gutin B, Litaker M, Islam S, Manos T, Smith C, Treiber F. Body-composition measurement in 9–11-year-old children by dual-energy X-ray absorptiometry, skinfold-thickness measurements, and bioimpedance analysis. Am J Clin Nutr 1996;63:287–92.
36. Deurenberg P, Pieters JJL, Hautvast JGAJ. The assessment of the body fat percentage by skinfold thickness measurements in childhood and young adolescence. Br J Nutr 1990;63:293–303.
37. Heyward V, Stolarczyk L. Applied body composition assessment. Champaign, Ill, USA: Human Kinetics, 1996.
38. NHES. National Health Examination Survey. Sample design and estimation procedures for a national health examination survey of children. In: National Center for Health Statistics Publication No. HRA 74-1005. Health Resources Administration: Rockville, Md, USA, 1973.
39. NCYFS. Summary of findings from National Children and Youth Fitness Study. Washington, DC: Department of Health and Human Services. 1985.
40. Sardinha LB, Teixeira PJ. Measuring adiposity and fat distribution in relation to health. In: Human Body Composition, 2nd ed. Champaign, Ill, USA: Human Kinetics, 2005.
41. Weststrate JA, Deurenberg P, van Tinteren H. Indices of body fat distribution and adiposity in Dutch children from birth to 18 years of age. Int J Obes 1989;13:465–77.
42. Seidell JC, Oosterlee A, Thijssen MAO, Burema J, Deurenberg P, Hautvast JGAJ, Ruijs JHJ. Assessment of intra-abdominal and subcutaneous abdominal fat: Relation between anthropometry and computed tomography. Am J Clin Nutr 1987;45:7–13.
43. Frisancho AR. New norms of upper limb fat and muscle areas for assessment of nutritional status. Am J Clin Nutr 1981;34:2540–5.
44. Lohman T, Roche A, Martorell R. Anthropometric Standardization Manual. Champaign, Ill, USA: Human Kinetics, 1988.
45. Gibson R. Anthropometric reference data. In: Gibson R. Principles of nutrition assessment. New York: Oxford University Press, 2005.
46. McCarthy HD, Jarrett KV, Crawley HF. The development of waist circumference percentiles in British children aged 5.0–16.9 years. Eur J Clin Nutr 2001;55:902–7.
47. Houtkooper LB, Lohman TG, Going SB, Howell WH. Why bioelectrical impedance analysis should be used for estimating adiposity. Am J Clin Nutr 1996;64(3 Suppl):436S–448S.
48. Schoeller DA, Luke A. Bioelectrical impedance analysis prediction equations differ between African Americans and Caucasians, but it is not clear why. Ann. N. Y. Acad. Sci. 2000;904:225–6.
49. Ward LC, Heitmann BL, Craig P, Stroud D, Azinge EC, Jebb SA, Cornish BH, Swinburn B, O'Dea K, Rowley K, McDermott R, Thomas BJ, Leonard D. Association between ethnicity, body mass index, and bioelectrical impedance: Implications for the population specificity of prediction equations. Ann NY Acad Sci 2000;904: 199–202.
50. Morrison JA, Guo SS, Specker B, Chumlea WC, Yanovski SZ, Yanovski JA. Assessing the body composition of 6–17-year-old black and white girls in field studies. Am J Hum Biol 2001;13:249–54.
51. Going S, Nichols J, Loftin M, Stewart D, Lohman T, Turri G, Ring K, Catellier D, Pickrel J, Blew R, Stevens J. Validation of bioelectrical impedance analysis (BIA) in Black, White and Hispanic adolescent girls. Int J Body Comp Res (in press).
52. Wagner DR, Heyward VH. Measures of body composition in blacks and whites: a comparative review. Am J Clin Nutr 2000;71:1392–402.
53. Ortiz O, Russell M, Daley TL, Baumgartner RN, Waki M, Lichtman S, Wang J, Pierson RN Jr, Heymsfield SB. Differences in skeletal muscle and bone mineral mass between black and white females and their relevance to estimates of body composition. Am J Clin Nutr 1992; 55:8–13.
54. Malina RM. Anthropometry. In: Maud PJ, Foster C (eds). Physiological assessment of human fitness. Champaign, Ill, USA: Human Kinetics Publishers, 205–19, 1995.
55. Malina RM. Variation in body composition associated with sex and ethnicity. In: Heymsfield SB, Lohman TG, Wang ZM, Going SB (eds). Human Body Composition, 2nd ed. Champaign, Ill, USA: Human Kinetics, 271–98, 2005.
56. Hampton MC, Huenemann RL, Shapiro LR, Mitchell BW, Behnke AR. A longitudinal study of gross body composition and body conformation and their association with food and activity in a teen-age population. Anthropometric evaluation of body build. Am J Clin Nutr 1966;19:422–35.
57. Trotter M, Hixon BB. Sequential changes in weight, density, and percentage ash weight of human skeletons from an early fetal period through old age. Anat Rec 1974; 179:1–18.
58. Chumlea WC, Guo SS, Kuczmarski RJ, Flegal KM, Johnson CL, Heymsfield SB, Lukaski HC, Friedl K, Hubbard VS. Body composition estimates from NHANES III bioelectrical impedance data. Int J Obes Relat Metab Disord 2002;26:1596–609.
59. Sun SS, Chumlea WC, Heymsfield SB, Lukaski HC, Schoeller D, Friedl K, Kuczmarski RJ, Flegal KM, Johnson CL, Hubbard VS. Development of bioelectrical impedance analysis prediction equations for body composition with the use of a multicomponent model for use in epidemiologic surveys. Am J Clin Nutr 2003;77:331–40.
60. Siervogel RM, Maynard LM, Wisemandle WA, Roche AF, Guo SS, Chumlea WC, Towne B. Annual changes in total body fat and fat-free mass in children from 8 to 18 years in relation to changes in body mass index. The Fels Longitudinal Study. Ann NY Acad Sci 2000;904: 420–3.
61. Cohn SH, Abesamis C, Zanzi I, Aloia JF, Yasumura S, Ellis KJ. Body elemental composition: Comparison between black and white adults. Am J Physiol 1977; 232:(4)E419–22.

62. Novotny R, Going S, Teegarden D, Van Loan M, McCabe G, McCabe L, Daida Y. 2005. Asian, Hispanic and White adolescent body size, composition and fat distribution. In: Exp Biol San Diego, CA: Federation of American Societies for Experimental Biology. 270.2:108.
63. Going S, Novotny R, McCabe G, McCabe L, Teegarden D, Lohman T, Van Loan M, Daida Y, Boushey C. 2005. Relationship of body mass index to percent fat in Asian, Hispanic and non-Hispanic adolescent girls. In: Exp Biol San Diego, Calif, USA: Federation of American Societies for Experimental Biology. 270.8:108.
64. Hattori K, Tatsumi N, Tanaka S. Assessment of body composition by using a new chart method. Am J Hum Biol 1997;9:573–8.
65. Wells JC. A critique of the expression of paediatric body composition data. Arch Dis Child 2001;85:67–72.
66. Maynard LM, Wisemandle W, Roche AF, Chumlea WC, Guo SS, Siervogel RM. Childhood body composition in relation to body mass index. Pediatr 2001;107:344–50.
67. Wells JC, Cole TJ. Adjustment of fat-free mass and fat mass for height in children aged 8 years. Int J Obes Relat Metab Disord 2002;26:947–52.
68. Wells JC, Coward WA, Cole TJ, Davies PS. The contribution of fat and fat-free tissue to body mass index in contemporary children and the reference child. Int J Obes Relat Metab Disord 2002;26:1323–8.
69. Horlick M, Wang J, Pierson RN Jr, Thornton JC. Prediction models for evaluation of total-body bone mass with dual-energy X-ray absorptiometry among children and adolescents. Pediatr 2004;114:e337–45.
70. Wang Z, Shen W, Withers RT, Heymsfield SB. Multicomponent molecular-level models of body composition analysis. In: Heymsfield SB, Lohman TG, Wang ZM, Going SB (eds). Human Body Composition, 2nd ed. Champaign, Ill, USA: Human Kinetics, 163–76, 2005.
71. Lohman TG, Going SB. Multicomponent models in body composition research: opportunities and pitfalls. Basic Life Sci. 1993;60:53–8.
72. Withers RT, Smith DA, Chatterton BE, Schultz CG, Gaffney RD. A comparison of four methods of estimating the body composition of male endurance athletes. Eur J Clin Nutr 1992;46:773–84.
73. Williams DP, Going SB, Lohman TG, Harsha DW, Srinivasan SR, Webber LS, Berenson GS. Body fatness and risk for elevated blood pressure, total cholesterol, and serum lipoprotein ratios in children and adolescents. Am J Pub Health 1992;82:358–63.
74. Williams DP, Going SB, Lohman TG, Hewitt MJ, Haber AE. Estimation of body fat from skinfold thicknesses in middle-aged and older men and women: A multiple component approach. Am J Hum Biol 1992;4:595–605.
75. Berenson GS, McMahon CA, Voors AW. Cardiovascular risk factors in children: The early natural history of atherscleros is and essential hypertension. New York: Oxford University Press, 1980.
76. Aristimuno GG, Foster TA, Voors AW, Srinivasan SR, Berenson GS. Influence of persistent obesity in children on cardiovascular risk factors. The Bogalusa Heart Study. Circulation 1984;69:895–904.
77. Berenson GS, Webber LS, Srinivasan SR, Voors AW, Harska DW, Dalferes ER. Bio-chemical and anthropometric determinants of serum B- and pre-B-lipoproteins in children. The Bogalusa Heart Study. Arteriosclerosis 1982;2:325–34.
78. Smoak CG, Burke DS, Freedman LS, Webber LS, Berensen GS. Relation of obesity to clustering of cardiovascular disease risk factors in children and young adults. The Bogalusa Heart Study. Am J Epidemiol 1987;125:364–72.
79. Voors AW, Harsha DW, Webber LS, Radhakrishnamurthy B, Srinivasan SR, Berenson GS. Clustering of anthropometric parameters, glucose tolerance, and serum lipids in children with high and low beta- and pre-beta-lipoproteins. The Bogalusa Heart Study. Arteriosclerosis 1982;2:346–55.
80. Sardinha LB, Going SB, Teixeira PJ, Lohman TG. Receiver operating characteristic analysis of body mass index, triceps skinfold thickness, and arm girth for obesity screening in children and adolescents. Am J Clin Nutr 1999;70:1090–5.

Development of an international growth standard for preadolescent and adolescent children: Reviewers

Acknowledgments

The editors of this supplement thank the following reviewers for their valuable participation in this project:

Tim Armstrong, World Health Organization
William C Chumlea, Wright State University School of Medicine
Philippe Froguel, Imperial College London
Ed Frongillo, Cornell University
Rolland Hauspie, Free University of Brussels
Reynaldo Martorell, The Rollins School of Public Health of Emory University
Kim F. Michaelsen, Royal Veterinary and Agricultural University
Luis A. Moreno Aznar, Universidad de Zaragoza
John Reilly, University of Glasgow
Lawrence M. Schell, State University of New York, Albany
Dennis M. Styne, University of California, Davis
Shumei S. Sun, Wright State University School of Medicine
Cesar Victora, Universidade Federal de Pelotas

We would like to thank those who provided administrative support for this project, including the staff of the United Nations University Food and Nutrition Programme at Cornell University and the United Nations University Center, Tokyo; and the World Health Organization.