The WHO Growth Chart: Historical Considerations and Current Scientific Issues

Mercedes de Onis\textsuperscript{a}, Ray Yip\textsuperscript{b}

\textsuperscript{a} Nutrition Unit, World Health Organization, Geneva, Switzerland;
\textsuperscript{b} Division of Nutrition, Centers for Disease Control, Atlanta, Ga., USA

Introduction

Growth assessment is the single measurement that best defines the health and nutritional status of children, just as it provides an indirect measurement of the quality of life of an entire population. Building from a history of pediatric practice, the premise for the use of growth as an indicator of health and nutritional status lies in the fact that ‘poor’ growth performance reflects, for the majority of the children who are so designated, deviations from favorable environmental conditions that support optimal growth and development in childhood.

The designation of a child as having impaired growth implies some means of comparison with a ‘reference’ child of the same age and sex. This need has made the choice of a normative growth reference an important issue that has received considerable attention in the last decades; especially as growth retardation during early childhood and its adverse implications for child development remains a disturbing problem worldwide [1]. This paper reviews the background and history of the currently used NCHS/WHO growth charts, discusses some of the contemporary scientific issues and outlines the recommendations made by a WHO Expert Committee held in November 1993 [2] related to the revision and development of a new international growth reference.
Background and History

For many years the pioneer data for children developed in the 1940s by Meredith [3] at Iowa were used mostly for individual assessment. These data were derived from a small and unrepresentative sample of US children, most of whom were of a high socioeconomic status. In the 1960s and 1970s and following on from the Iowa data, two data sets were frequently used as growth references: the Harvard growth curves [4] and the Tanner growth curves from the UK [5]. In 1966, a simplified combined-sexes version of the Harvard growth curves was widely disseminated by the World Health Organization [6] and effectively established the Harvard curves as the international growth reference. This data, derived from the growth of Caucasian children in Boston from 1930 to 1956, offered the advantages of having been compiled longitudinally, of being widely available in Nelson’s Textbook of Pediatrics [4], and of being used already by pediatricians in many countries. The major criticisms of the Harvard data come from the small numbers of children included, the limited genetic representativeness, and outdatedness.

During the next decade, as a result of the limitations of the Harvard and other available reference data and the desire in the US for a more contemporary reference that better represented the population of that country [7], the US National Academy of Sciences recommended in 1974 a new set of anthropometric data as the US national growth reference [8, 9]. A combined National Center for Health Statistics (NCHS) and Centers for Disease Control (CDC) task force, using cross-sectional data from the US Health Examination Surveys of NCHS and longitudinal data from the Fels Research Institute, constructed a set of smoothed percentile distributions for attained weight, height, and head circumference from birth to 18 years [10, 11]. A description of these newer charts and their limitations can be found elsewhere [12, 13].

In 1975, a working group was convened to advise WHO on the use of anthropometric indicators of nutritional status in surveys and for nutritional surveillance. As a result, recommendations were made on the use of children’s height and weight data which included the use of a reference population for international comparison [14]. Among other criteria, the sample was to be from a cross-sectional study of a well-nourished population and was to include at least 200 children of each sex in every age group. The sampling methods were to be well defined and reproducible and the anthropometric measurements were to have been made by trained staff using well-designed, calibrated equipment. The anthropometric measurements were to include a selection of measurements sufficient to evaluate the nutritional status of children. Finally, the anthropometric data were to be available for anyone wishing to use them and
the procedures used for smoothing the curves were to be adequately described and documented [14].

Although none of the three anthropometric data sets available at that time (the measurements of Dutch children reported by Van Wierringen [15]; those of US National Academy of Sciences [9]; and those of British children reported by Tanner et al. [5]) met all the criteria recommended by the WHO working group, the NCHS/CDC growth reference was identified as being the most suitable for use as the international growth reference [14, 16, 17].

The NCHS/CDC growth curves were developed by combining two distinct data sets representing different age groups compiled over different decades. The reference for ages 0–23 months is based upon a group of children in the Ohio Fels Research Institute Longitudinal Study from 1929 to 1975. This reference reflects the growth of children who were fed primarily with infant formula and who were of restricted genetic, geographic, and socioeconomic background. The height curves for this part of the reference were based on recumbent length or supine measurements (the Fels curves). The reference from 2 to 18 years was based on data of three cross-sectional US representative surveys conducted from 1960 to 1975. The height curves for this part of the reference were based on standing height measurements (the NCHS curves). It was intended that length measurements be interpreted with the Fels curves, and height measurements with the NCHS curves. One major limitation of this reference is therefore that it is made up from two unrelated samples or sets of curves. Ideally, a reference should be based on a single set of curves or on a single survey sample.

In 1980, a mainframe computer software version of the reference was developed by the Centers for Disease Control to facilitate the interpretation of growth data from surveys or clinical studies. In order to formulate the software-based reference, the original height or weight distributions were slightly modified by a normalization procedure [12]. Throughout the 1980s, several microcomputer-based software versions of the NCHS-WHO reference were developed and supported by CDC and WHO [18]. These software-based references have contributed to the wide acceptance of the concept of the international reference because they simplified the handling of anthropometric data from surveys, surveillance, and clinical studies.

**Issues Related to Reference Selection and Application**

*Reference versus Standard*

On theoretical grounds it is important to distinguish between a reference and a standard [19]. A reference is defined as a tool for grouping and analyzing
data and provides a common basis for comparing populations without making inferences about the meaning of observed differences. A standard, on the other hand, embraces the notion of a norm or desirable target, and thus involves a value judgement. However, because reference data embody certain characteristics or patterns of normality, they have been widely used to make inferences about the health and/or nutrition of individuals and populations; that is, they have been treated as optimum targets, or standards, and any deviations have been assumed to have a fixed and particular meaning. Much of the justification for this is provided by extensive evidence that, in populations, the effect of genetic differences on the growth of children is small in comparison with the large differences observed due to environmental factors.

Therefore, recognizing that in practice it is almost impossible to prevent the use of reference data as standards for judging the nutritional status of individuals and populations, it is recommended that care should always be taken to choose references that resemble, as far as possible, true standards, so that the same deviation from the reference data has the same biological meaning. Beyond the consideration of choosing a reference population, efforts should be concentrated on the appropriate use of the reference as a general guide for screening and monitoring, and not as an absolute criterion to define ‘malnutrition’ or pathology.

Local versus International Reference Data

An international reference is clearly needed to compare the nutritional status of populations in different parts of the world. There is evidence that the growth in height and weight of well-fed, healthy children, or children who experience unconstrained growth, from different ethnic backgrounds and different continents is reasonably similar at least up to 5 years of age [20, 21]. It is accepted that there is some variation in the growth patterns among children of different race or ethnic groups in developed countries; however, these variations are relatively minor compared to the large worldwide variation in growth related to health, nutrition and socioeconomic status [22, 23]. To illustrate this point using two very different populations, figure 1 shows how the growth performance of Indian children not subject to socioeconomic and dietary constrains, nearly corresponds to the NCHS/WHO international reference values. The same pattern of growth is seen among affluent children in seven different cities of India [24]. Similarly, in Brazil, children in the highest quartile of per capita income have a height-for-age Z-score distribution that overlaps with that of the international reference [25]. Figure 2 illustrates how the entire height-for-age Z-score distribution for Brazilian children shifts to the right in moving from the lowest to the highest income quartile distribution. Several other affluent populations from different ethnic backgrounds have
been shown to have a growth pattern similar to the international reference. For this reason, the use of a common reference has the advantage of uniform application allowing international comparisons without losing the usefulness for local application. Such advantages outweigh the disadvantage of not taking into account minor racial and ethnic variations.

Beyond lacking value for international comparisons, there are also several reasons for not developing a local reference or standard [26]: (1) many populations in less-developed areas experience growth deficits as a result of poor health and nutrition, and therefore the reference developed from such populations has less screening value for the detection of health and nutritional disorders; (2) significant secular changes in growth status within a relatively short period of time may render a local reference less useful for clinical screening; (3) proper reference development is not a task that can be done easily and frequently, and (4) it is very costly to develop local references.

Probably more important than which reference is chosen is the way in which the reference is interpreted, and the clinical and public health decisions that will be based upon it. Criteria for selection of cut-off points used in the assessment of ‘high-risk’ individuals and populations should be based on evidence of increased risk for mortality, morbidity or impaired performance. Future research should therefore attempt to identify the range of child growth associated with optimal long-term health outcomes, as well as ranges associated with specific adverse outcomes. These results may lead to changes in the traditionally used cut-off points based on statistical criteria.

Fig. 1. Comparison of the 50th percentile of heights of affluent Indian girls (Ludhiana) with NCHS/WHO reference values. Source: Agarwal et al. [24].
**Factors Affecting the Use and Interpretation of Growth References**

A number of nonpathological factors may influence growth during childhood. Some of these factors have the potential to alter the interpretation of growth status at both the individual and population level.

*Feeding Practice.* Existing data in the published literature suggest that the mode of feeding during infancy significantly influences growth patterns. Breast-fed infants living under favourable conditions and studied in various geographical areas have been reported to follow negative trends relative to the NCHS-WHO weight-for-age, and possibly length-for-age, percentiles during the latter half of the recommended exclusive breastfeeding period [27, 28].

*Racial and Ethnic Variation.* There is evidence that the following variations exist even though relatively minor in contrast to variations related to socioeconomic status: (1) weight-for-height status is lower for children of the Indian subcontinent [29]; (2) Hispanic children in Central and South America appear to have higher weight-for-height status [29, 30]; (3) Black children have lower birth-weight and smaller size during infancy, but exceed the size of white children after 2–3 years of age [31].

*Sex.* In the age range of concern (birth to 10 years) the average weights and heights of boys are consistently larger than those of girls. Gender-specific
references are therefore recommended, with the possible exception of disaster situations where simplification of procedures may require the use of unisex references.

**Size at Birth.** Birth-size and intrauterine growth status appear to be strong determinants of later childhood growth status even in the range commonly accepted as normal birth weight. Intrauterine growth-retarded infants grow more slowly than preterm infants of the same birth-weight [32].

**Parental Stature.** Parental size, especially height, is a determinant of both birth-weight and later childhood growth.

**Altitude.** The environmental factor that has most effect on growth but is not associated with socioeconomic factors is altitude (or oxygen partial pressure); high altitude leads to reduced birth weight and lesser subsequent growth [33].

With the exception of sex and possibly of infant feeding mode, most of the factors discussed here do not seem to warrant a separate reference to account for the observed differences. Rather, awareness of these factors can help to adjust for possible variations when comparisons are made across groups.

Regardless of the approach taken, the definition of the reference population must include consideration of variability. A reference with an inappropriately small variability will result in more children classified as being of clinical concern as compared to a reference with more variability. Since references are intended for international use, they should reflect the variability observed cross-nationally in well-nourished, healthy populations, including at least partly variability due to genetic differences.

**The Growth of Breast-Fed Infants**

As part of the preparatory work for the recently held Expert Committee on the use and interpretation of anthropometry [2], WHO established a working group on infant growth to examine the growth of breast-fed infants living under favourable environmental conditions and see whether they differed substantially from accepted international references; and whether those differences were of practical importance to clinical practice and public health policy.

Concern had been expressed in the literature that breast-fed infants, living under favorable conditions, were growing less well than expected when compared to the NCHS-WHO growth reference. The negative deviations in weight-for-age appeared large enough to cause health workers to draw inappropriate conclusions about the growth of these infants. This possibility was particularly worrisome in environments where breast-feeding is key to survival. Alternatively, true growth failure in a breast-fed infant may have been misinterpreted.
as the 'normal negative deviation' in growth expected when breast-fed infants are compared to the current international reference.

A survey of investigators with data sets on growth of breast-fed infants was conducted. Seven data sets fulfilled the criteria for inclusion in the initial review and were examined in detail: one each from Canada, Denmark, Finland, Sweden, the UK, and two from the US. Of the 453 infants followed in the seven studies, 226 were breast-fed for at least 12 months and not given solid food, formula or other milks before 4 months of age. These infants were referred to as the 'breast-fed set'. Maternal education levels were high in the studies reporting this variable, and mean birth-weight in the seven studies was above 3,400 g. A complete description of these analyses is available elsewhere [27, 28].

Figure 3 presents the Z-scores patterns of the 'breast-fed set' infants relative to the NCHS/WHO reference values. Mean weight-for-age declined continuously from 2 to 12 months to a low of almost -0.6 SD at 12 months. The magnitude of the decline in length-for-age was not as great, with the mean Z-score tending to stabilize or increase after 8 months; the mean value at 12 months was approximately -0.3 SD. Mean weight-for-length at 12 months was also below the NCHS-WHO reference median (approximately -0.3 SD). The declines in Z-scores observed during the period of complementation are difficult to interpret. One possibility is that these are artifacts due to technical problems in the construction of the NCHS-WHO references [2]. The declines may also be a consequence of specific weaning practices within the population studied, or the result of other physiological effects attributable to continued
Table 1. Comparison of the data sets used in the NCHS/WHO growth reference

<table>
<thead>
<tr>
<th>Data source</th>
<th>Representative?</th>
<th>Age range years</th>
<th>Stature measurement</th>
<th>Study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fels</td>
<td>no</td>
<td>0.0–2.99</td>
<td>recumbent</td>
<td>longitudinal</td>
</tr>
<tr>
<td>NCHS</td>
<td>yes</td>
<td>2.0–17.99</td>
<td>standing</td>
<td>cross-sectional</td>
</tr>
</tbody>
</table>

Source: Gorstein et al. [34].

1 Representative of the entire US population.

breast-feeding, or even due to unidentified characteristics in the populations used to construct the NCHS-WHO reference and the ‘breast-fed set’.

Based on this review [27, 28], the working group concluded that infants fed according to WHO recommendations and living under conditions favoring the attainment of genetic growth potential grew less rapidly and deviated significantly from the reference. Given the short- and long-term consequences of growth failure, the dangers both of the premature introduction of complementary foods and their undue delay – described as the ‘weanling’s dilemma’ – the working group recommended the development of a new weight and length reference which would enhance the nutritional management of infants. The specific issues to take into consideration in the development of such a reference and the characteristics of the sample to be selected as a reference population have been described elsewhere [2].

Of special concern is the role of complementary foods in determining growth patterns. Research needs to be done in advantaged populations to examine the effects of the quality of complementary foods on growth and other health outcomes from 4 months onwards. The lack of this information makes it difficult both to evaluate growth in the latter half of infancy, and to identify the etiology and measure the rates of stunting and/or wasting in infants within and between populations.

Current Problems of the NCHS/WHO International Growth Reference

*Height-for-Age Disjunction*

As mentioned earlier, one major limitation of the current international reference is that it is made up from two unrelated sets of curves (table 1). As a result, there is a significant disjunction in the height curve between the length-based
Fig 4. Mean height-for-age Z-scores across age for four ethnic groups of low-income US children monitored by the CDC Pediatric Nutrition Surveillance System, illustrating the marked shift in mean Z-scores across the 24-month disjunction. From Yip et al. [31].

Fels curves for children under 2 years of age and the height-based NCHS curves for older children. This marked discrepancy in estimated height status immediately before and after 24 months of age, where the two sets of data merge, is responsible for the abrupt change consistently noted at 2 years of age in the mean Z-score of height-for-age and the prevalence of low height-for-age when analysing growth data from large surveys. A detailed description of the nature and magnitude of the height-for-age disjunction can be found elsewhere [2, 34].

Figure 4 uses the mean height-for-age Z-score across age for four racial groups of low-income US children monitored by CDC to illustrate the significant changes described above [31]. The impact of the 24-month disjunction on the prevalence of low height-for-age for these children is illustrated in figure 5. The magnitude of these deviations warrants caution when using the current reference to interpret the growth status of children covering a wide range of ages, as in surveys or in the context of growth monitoring.

**Upward Skewness of Reference Population**

The height distributions of children of a given age are normally (symmetrically) distributed. On the other hand, the distributions of weight-for-
Fig 5. Change in prevalence of low height-for-age of low-income US children monitored by the CDC Pediatric Nutrition Surveillance System, illustrating the impact of the 24-month disjunction. Source: WHO [2].

age and weight-for-height are normal in some developed countries but are skewed towards the higher end in others. In fact, the current NCHS-WHO reference based on United States’ children is markedly skewed, reflecting a substantial level of childhood obesity. Using such a reference as a ‘standard’ for optimal growth or health is unwarranted since the upward skewness may reflect an ‘unhealthy’ characteristic of the sample. In addition, given the secular trends of increased childhood and adult overweight observed in the USA and in some other populations, it is of concern that, unless this issue is dealt with, future references will tend to be further and further skewed to the right, thereby resulting in the misclassification of overweight children as ‘normal’.

One potential solution is to use the standard deviation of the lower half of the distribution to model the upper half. The main justification for this suggestion is that variations in the lower half of the weight distributions among developed countries (non-constrained populations) are relatively small. However, at present it is felt that not enough is known about the normality of weight distribution in children of different ages so that the issue of skewness should be the subject for further research [2].

Revision Effort of the Current NCHS-WHO Growth Reference

Because of the potential for misinterpretation of anthropometric data resulting from the above-mentioned problems of the current reference, a revision was carried out by CDC in 1990 to correct some of the irregularities. The revision added additional data from the original US national
Fig. 6. Height-for-age disjunction at 24 months of the current NCHS-WHO reference and of the revised NCHS reference. Source: WHO [2].

survey to expand the sample size for more reliable curve formulation, as well as to reduce the use of the Fels sample for 12–24 months. In addition, different statistical procedures were developed to ensure greater accuracy of the formulated curve. Figure 6 shows that the gap of height values between the Fels curves and the NCHS curves was reduced from the original 2.5 cm to 0.5 cm after the revision. It should be noted that this reflects only the difference between recumbent length and standing height, which in this case was small since the children were properly stretched for the measurement.

Since a national United States representative sample of children under 12 months of age was not available, the Fels sample was retained for this age group. However, a new procedure based on modelling the individual growth pattern was used to overcome the inadequate measurement points for the sample. Unlike the original reference, where the Fels and the NCHS curves were kept as separate sets, the revised reference merged them at 12 months of age to create a single set, as at this age an adequate fit was observed.

The adoption of the revised curves as the updated international reference for general use is not recommended for several reasons [2]. Rather, the formulation of a truly international reference based on well-conducted surveys covering broad populations from several countries might be more acceptable than data from a single country. Some of the desirable characteristics for the anthropometric data to be used in the development of such a reference population are listed in table 2.
Table 2. Desirable characteristics for anthropometric data to be used in the development of an international growth reference

Several countries from different geographical regions should be included
Based on healthy populations with unconstrained growth (not necessarily representative)
Adequate sample sizes and procedures
The raw data should be available
Age range from birth to adolescence
Standardized quality control and measurements with documentation of standardization procedures
For adolescents, measures of sexual maturity should be available
Secular trends in growth should be small or absent

Adapted from: WHO [2].

**Growth Velocity Curves**

Growth velocity curves will detect growth faltering earlier than attained growth charts. Velocity (rate of growth) represents what is happening now, whereas attained growth represents the sum of all that has happened in the past [35]. However, such curves are more difficult to interpret and require repeated visits at regular intervals. In addition, growth over short periods of time show high variability, a finding which along with the low precision of some measurements will limit the ability to detect growth changes. For proper decision-making, it is important to use time intervals for which the expected growth is greater than the combined errors of two repeated measures.

The significance of a change in weight or height clearly depends on the time over which that change is observed. Failure to gain weight over 3 days has a different significance from failure to gain weight over 3 months. Therefore, when velocities are reported, it is important to indicate the time over which they are measured. Different time intervals will be appropriate in different situations. They will depend on, among other things, the age of the children and whether weight or height gain is the main object of interest. The smaller time intervals will have the greatest error. The effect of the wide variability in normal growth rates on the use and interpretation of growth velocity reference data is also an important research issue.

One important difficulty encountered in constructing longitudinal reference data is the sample size required for the accurate location of the extreme percentiles. This difficulty can be overcome by constructing curves the shape
of which are based on longitudinal data, but whose amplitudes (in the sense of distance between the centiles) are based on a large-scale cross-sectional survey [36]. This was the method used in creating one of the longitudinal standards presently available [37].

Accurate estimates of velocity also require that the same children should be examined at defined intervals. Longitudinal studies of this kind will usually not be feasible. Even if they are, another important issue concerns the frequency of data collection. Should data be collected once a month, once every 3 months, or using any other time interval? For length and weight during infancy, intervals between 2 weeks and 3 months have been proposed [38]. In available velocity standards the data points are at intervals of 3 months or longer. If a falling off in velocity is to provide early warning of growth failure, it must be detected over much shorter periods than 3 months. The expected velocity over any given month can be obtained by interpolation, but the SD of 3-month gains cannot be applied to gains over 1 month. The shorter the interval, the larger the SD [39]. There is no doubt that clinical practice requires velocity reference data; however, considering the difficulties involved in constructing growth velocity curves, the justification for their general public health use would have to be carefully evaluated before embarking on this task.

Conclusions and Recommendations of the WHO Expert Committee Related to the Revision and Development of an International Reference

The Expert Committee on ‘Physical status: the use and interpretation of anthropometry’ reaffirmed the previous WHO position of using a single international reference. However, because of the significant technical drawbacks of the current NCHS-WHO reference population, especially for population-based application, an update or replacement in the near future was recommended [2]. Updating a reference, or developing an entirely new one, is an extremely complex, costly and time-consuming undertaking. Such a daunting task cannot be treated lightly. To be of lasting value, therefore, it is clear that the next reference must be exceptionally well prepared. The Expert Committee identified some of the desirable characteristics for anthropometric data to be used in the development of a new growth reference. Applying the Committee’s recommendation that the reference be based on data from numerous countries will help prevent both the technical and political difficulties that have arisen from using a single country’s child-growth pattern as a worldwide ‘standard’.

In addition to developing a more acceptable international reference, efforts should be concentrated on the appropriate use of the reference. The reference
should be used as a general guide for screening and monitoring and not as a fixed standard that can be applied in a rigid fashion to children from different ethnic, socioeconomic, and nutritional and health backgrounds.

For clinical or individual-based application, the NCHS/WHO growth curves should be used as a screening tool to detect children at greater risk of health or nutritional disorders; and they should not be viewed as a diagnostic label for ‘malnutrition’.

For population-based application, the reference population can be used adequately for cross-comparison and monitoring purposes. In a given population or community, a high prevalence of anthropometric deficit will be indicative of significant health and nutritional problems; however, it is not only those children below the cut-off point who are at risk, the entire population is at risk, and the cut-off point should be used only to facilitate the application of the indicator.

References


Dr. Mercedes de Onis, World Health Organization, Nutrition Unit, 20, Av. Appia, CH–1211 Geneva 27 (Switzerland)