

THE USE AND CONSTRUCTION OF ANTHROPOMETRIC GROWTH REFERENCE STANDARDS

T. J. COLE

MRC Dunn Nutrition Centre, Milton Road, Cambridge CB4 1XJ

CONTENTS

USE OF GROWTH STANDARDS	20
HISTORY	20
DEFINITIONS	21
REFERENCE OR STANDARD	21
LOCAL OR INTERNATIONAL	22
SCALES OF MEASUREMENT	22
<i>Centile</i>	23
<i>Percentage of median</i>	23
<i>SD score</i>	23
DISTANCE AND VELOCITY STANDARDS	26
<i>Weight for height</i>	28
MARGINAL AND CONDITIONAL STANDARDS	28
<i>Marginal or unconditional</i>	28
<i>Conditional</i>	29
PURPOSE OF STANDARDS	31
<i>Education</i>	31
<i>Screening</i>	31
<i>Clinical</i>	32
<i>Group summary</i>	32
EFFECT OF AN INAPPROPRIATE STANDARD	32
FUTURE DEVELOPMENTS USING SD SCORES	34
<i>Display of SD scores</i>	34
<i>Indices based on SD scores</i>	34
GROWTH STANDARD CONSTRUCTION	36
PRINCIPLES	36
GROWTH CURVES	36
<i>Smoothing</i>	36
<i>Infancy models</i>	37
<i>All childhood models</i>	38
CENTILE SPECIFICATION	39
<i>Normal distribution</i>	40

<i>Transformed normal distribution</i>	41
<i>Constrained non-parametric</i>	45
<i>Unconstrained non-parametric</i>	45
<i>Measurement error</i>	46
CONCLUSIONS	46
USE OF STANDARDS	46
CONSTRUCTION OF STANDARDS	46
REFERENCES	47

USE OF GROWTH STANDARDS

HISTORY

Just over a century ago Bowditch (1891) published the first account of growth standards, using curves based on Galton's percentiles. Galton had invented anthropometric percentiles, as he called them, to summarize the distribution of body measurements of '9337 persons measured in my Anthropometric Laboratory at the [1884] International Health Exhibition' (Galton, 1885). He compared the distributions of measurements for the two sexes by identifying the percentile for each sex where the distributions crossed, counting up for the men and down for the women. Thus in 'strength of squeeze', for example, he observed that the 7th percentile for men was the same as the 93rd percentile for women.

Bowditch (1891) extended this idea by displaying percentiles of height for Massachusetts children on a chart, so they appeared as curves plotted against age. By putting age on the abscissa he allowed changes in size with age, i.e. notional growth, to be displayed at the same time.

During the first half of the twentieth century, the use of growth charts expanded considerably. Various sets of reference data for height and weight were developed, most notably the Harvard standards (Vaughan & McKay, 1975). More recently other types of chart have been devised, for example the clinical longitudinal chart of Tanner (Tanner *et al.* 1966), the American NCHS chart, adopted for international use by WHO (Hamill *et al.* 1977; Lavoie-Pierre *et al.* 1983), and the conditional charts of Cameron (1980). At the same time other anthropometric measurements have been investigated, for example body circumferences and skinfold thicknesses. Tanner (1981) provides a comprehensive history of the development of growth studies and growth standards.

The methods of making standards, i.e. converting reference data to smooth percentile curves, have also changed enormously in recent years, with the advent of fast computers and improved statistical techniques. Originally the curves were drawn by eye, but in the last five years there has been something of an explosion in the statistical literature on the subject.

The clinical value of displaying anthropometric status and growth together with reference curves on a growth chart has been repeatedly demonstrated. Ironically, though, there remains widespread misunderstanding, even now, as to how precisely the growth chart should be interpreted.

The purpose of this review is twofold: to summarize how growth charts, and the reference standards that underlie them, are currently used and constructed, and to highlight ways in which their use might be extended in the future.

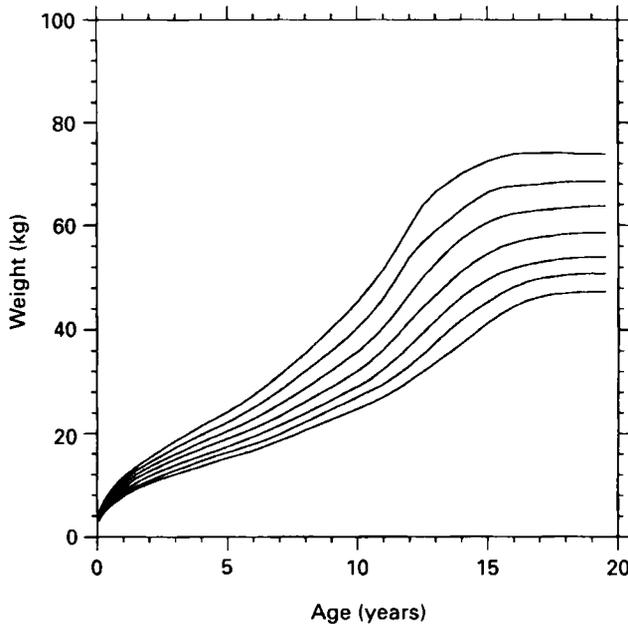


Fig. 1. Centile chart for Dutch girls' weight in 1980 (Roede & Van Wieringen, 1985), showing seven centiles from the 3rd to the 97th.

DEFINITIONS

A growth standard or growth reference is a dataset representing the distribution of a given anthropometric measurement as it changes with some covariate – usually age – in the two sexes, based on a specified reference sample of children. The distribution is usually summarized by selected centiles including the median (50th centile), and the mean and standard deviation (or SD) may also be given. Note that the term 'centile' is synonymous with Galton's 'percentile'. The set of centiles chosen to define the standard is conventionally symmetric about the median, with up to seven distinct centiles, and usually including extreme values such as the 3rd and 97th or the 5th and 95th. Centiles corresponding to $-2SD$ or $-3SD$ are also sometimes used.

Thus a growth standard consists of several smooth centile curves, which when plotted out make up a growth chart, with the relevant anthropometric measure on the ordinate and age (usually) on the abscissa. As an example, Fig. 1 shows a growth chart for weight in Dutch girls (Roede & Van Wieringen, 1985). Sometimes another anthropometric measure is on the abscissa, e.g. height on a weight-for-height chart.

REFERENCE OR STANDARD

The term 'growth standard' is ambiguous, and this ambiguity has generated confusion and controversy over the years. The issue is to do with the quality of growth that the standard represents. Does it represent 'good' growth, i.e. a norm to aspire to, or is it 'reference growth', a yardstick to be used for comparison purposes (WHO, 1986)? This question is not academic – a growth standard based on Western children and used in a developing country represents a size of child seen only in the most well-off sections of society. Clinically such a standard is unhelpful for the majority of children, and politically it may be unwelcome if it implies that their small size is due to poverty. Against this, there are

situations where growth clearly is inadequate, and the standard then provides a reminder of what might be possible in better conditions.

Using a growth standard as a norm implies that it represents optimal growth. This is a dubious assumption – a larger child is not necessarily a healthier child, and with obesity the reverse is true. In any case, even within the developed world there are clear differences in adult height. Young men in The Netherlands for example are more than 5 cm taller on average than young men from the USA (Hamill *et al.* 1977; Roede & Van Wieringen, 1985; Tanner & Davies, 1985). This is not easy to explain in health terms.

To avoid arid arguments about the nature of healthy growth, the second definition is generally preferred – a reference standard provides a reference, not a norm. In practice this means that the standard should not be used in isolation to reach a clinical decision. To reduce ambiguity some authorities recommend use of the phrase ‘growth reference’, thus avoiding ‘standard’ altogether (WHO, 1986). This course is not followed consistently here, but the terms ‘growth standard’ and ‘growth reference’ should be viewed as synonymous.

LOCAL OR INTERNATIONAL

A distinct but related issue concerns the use of international as opposed to locally based standards. The advantage of an international standard, such as the NCHS standard adopted by WHO (Hamill *et al.* 1977; Lavoie-Pierre *et al.* 1983) is that it simplifies comparisons between regions or countries – they all have a common reference. The main disadvantage is that the pattern of growth shown by the standard may be quite inappropriate in particular regions of the world, for example India, where the children are considerably smaller than in the USA. Against this, some authors (Goldstein & Tanner, 1980; Graitcer & Gentry, 1981) argue that if elite children in the region can achieve growth matching that of the standard, then the standard serves as a norm representing achievable growth for those who are less well-off. This then brings us back to the definition of the standard as norm.

The short answer to the question ‘International or Local?’ is ‘It depends’. Where international comparisons are required, an international standard simplifies the collection and classification of anthropometry. Conversely if clinical decisions are needed, or if a statistical analysis is used to adjust anthropometry for age, then a local standard is probably more appropriate.

In practice, most Western countries have developed their own national standards, which are used as local norms for clinical purposes. Another class of growth standards applies to genetically unusual groups of children, where the pattern of growth differs significantly from the local norm, for example Down’s Syndrome (Cronk *et al.* 1988) or Turner’s Syndrome (Ranke *et al.* 1983; Lyon *et al.* 1985; Naeraa & Nielsen, 1990). In such cases there is no value in using an international standard, as the differing growth performance is due to the child’s genotype, and international comparisons are better made using the syndrome-specific standard.

SCALES OF MEASUREMENT

Growth standards are used to compare individuals or groups relative to other children of the same age and sex. Growth monitoring relies on two simple principles: children growing normally are more likely to be found in the body than the tails of the anthropometric distribution, and when followed up they tend to maintain their position in the distribution. In detail, the process of growth is not as smooth as the use of a growth chart implies – during the first year at least, growth in individuals proceeds in fits and starts

(Lampl *et al.* 1992). However, for periods of time exceeding a few weeks the discrete nature of the growth curve is averaged out, so that the assumption of a smooth curve is not unrealistic.

Thus when anthropometry is age-sex adjusted and plotted against age, the expected pattern of growth is seen as a horizontal line on the chart. This form of presentation, discussed in detail in a later section, makes abnormal growth easier to detect. There are three ways to adjust anthropometry for age and sex.

Centile

The growth chart displays anthropometric measurements in terms of centiles. Centiles indicate the proportions of the distribution that lie below them; 50% of the reference population lies below the median (or 50th centile), and 3% below the 3rd centile. A child's value plotted on the chart can, with reasonable accuracy, be converted to a centile by interpolating between adjacent centile curves. This is fine so long as the value is between the highest and lowest curves. However, if it falls above the highest or below the lowest, then the conversion is inaccurate – the centile scale is bounded between 1 and 99, and in the tails of the distribution the centiles are very widely spaced. As a result, children with a growth deficit who fall below the 3rd or 5th centile, including many in the developing world, have only poorly specified centiles. This is a serious deficiency of the centile chart, which has led to the use of other scales of measurement.

Against this, an important advantage of centiles is that they can be derived whatever the form of the distribution, even for measurements with frequency distributions of unusual or unknown shape.

Percentage of median

A simple and widely used alternative to classification by centile is to express data as a percentage of the median for age and sex. For example, the protein-energy malnutrition (PEM) classification scheme of Waterlow (1972) is based on % weight-for-age, defined as

$$\left(\frac{\text{weight}}{\text{median weight for age and sex}} \right) \times 100. \quad (1)$$

This also applies to supine length and height, and it avoids the problems that centiles pose in the extremes of the distribution. In addition it is easy to calculate and to understand.

The major disadvantage of percentage-of-the-median is the interpretation of a given deficit – it depends on both the measurement and the age when it is measured. Fig. 2 shows the Dutch girls' weight centiles of Fig. 1 expressed as percentages of the median. The 25th centile is about 92% of the median until puberty, when it falls to 88%. The corresponding centile for height in early life (not shown) is higher, about 97% of the median. Thus a given deficit in height-for-age is more extreme in centile terms than the same deficit in weight-for-age. Also, both deficits are more serious at 9 months of age than at 12 years.

A cutoff of 120% of the median for Body Mass Index (BMI; weight/height²) is often used to define obesity, and this suffers from the same problem. Waterlow *et al.* (1977) urged that centiles and percentage-of-the-median should not be used; instead they recommended the use of SD scores.

SD score

In certain special cases where the underlying distribution is known, centiles can be expressed in terms of the mean and standard deviation of the distribution. This has important advantages in terms of the precision of the estimated centiles (Healy, 1974). By far the most common example is the normal (or Gaussian) distribution, where because it

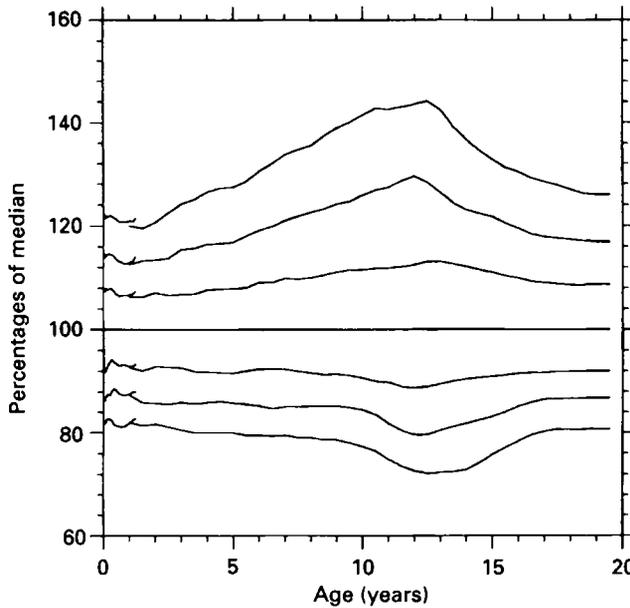


Fig. 2. The Dutch girls' weight centiles of Fig. 1 redrawn as percentages of the median. The median appears as a straight line at 100%.

is symmetric, the mean and median are the same. Thus a given centile is defined simply by its distance, in standard deviation units, from the mean. For example, the 97th centile is 1.88 SD above the mean, while the 3rd centile is the same distance below the mean. Normal distribution tables provide the conversions for other centiles. Height is known to have a distribution close to normal for most of childhood.

The SD equivalent to each centile is known as its SD score, or Z score or SDS. SD scores are by definition normally distributed, with a mean of 0 and an SD of 1. The SD score for height is defined as:

$$\left(\frac{\text{height} - \text{median height for age and sex}}{\text{SD of height for age and sex}} \right) \tag{2}$$

where the median and the SD are obtained from a published standard. SD scores are interchangeable with centiles wherever the underlying distribution is known, and they avoid the boundary problems of centiles – they can be as large or as small as necessary. Thus they represent the most appropriate scale of measurement for anthropometry, and their only disadvantage is that they cannot easily be applied to measurements that are not normally distributed.

The relationship between SD score and percentage-of-the-median can be illustrated by combining equations (1) and (2). Divide equation (2) by median height and multiply by 100 to give:

$$\left(\frac{100[\text{height}/\text{median height for age and sex}] - 100}{100[\text{SD of height for age and sex}/\text{median height for age and sex}]} \right)$$

or

$$\left(\frac{(\% \text{height-for-age} - 100)}{\% \text{CV}} \right) \tag{3}$$

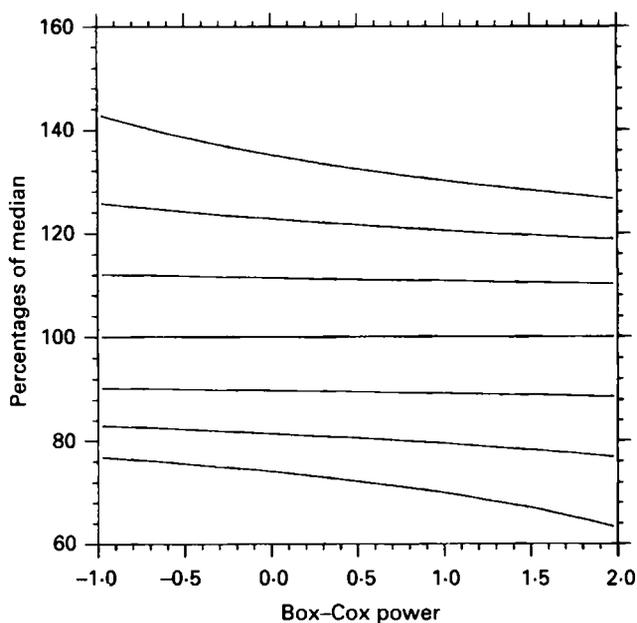


Fig. 3. Seven centiles, from the 3rd to the 97th, drawn as percentages of the median with varying degrees of skewness, showing the Box-Cox power transformation needed to make them normally distributed. The chart is based on equation (5), using a CV of 16%.

where %CV is the percentage coefficient of variation of height for age and sex. The CV is usually defined as SD/mean, but here the mean and median are the same, so either can be used. Equation (3) shows that if the %CV is a constant, the SD score and percentage-of-the-median are exactly equivalent.

Measurements that are not normally distributed can be transformed to make them closer to normal. A logarithmic transformation is often used for weight to reduce the skewness. Here the SD score is again defined in terms of the median and CV of the measurement (Cole, 1988):

$$\frac{100}{\%CV} \log_e \left[\frac{\% \text{weight-for-age}}{100} \right]. \quad (4)$$

A more general approach to the problem of non-normal distributions is to use a power transformation (Box & Cox, 1964). The measurement is raised to some power λ , which is chosen to remove the skewness in the distribution. If $\lambda = 1$ this is equivalent to using the untransformed measurement, while $\lambda = 0$ is the same as the log transform. Equation (3) can be extended to obtain SD scores in the general case, again illustrated with weight (Cole & Green, 1992):

$$\left(\frac{100 [\% \text{ weight-for-age}/100]^\lambda - 100}{\lambda \%CV} \right). \quad (5)$$

Substituting $\lambda = 1$ in (5) converts it back to (3). Fig. 3 shows how the centile spacings change, as percentages of the median, when the power λ varies over the range -1 to $+2$. When λ is 1 the centiles are symmetric; smaller values introduce progressively more positive skewness, while larger values provide negative skewness.

So this approach provides a convenient way of handling measurements whose

distributions are neither normal nor log-normal. It also provides a flexible framework for constructing growth standards (see later sections).

DISTANCE AND VELOCITY STANDARDS

The previous section has shown how to quantify a child's position in the distribution as a centile or SD score. Usually this is a one-off measure, based on a single measurement, which gives no clues as to the growth pattern that has led the child to its current position. It indicates only how extreme the child is in terms of current size or status.

If two successive measurements for the same child are available they are unlikely, in centile terms, to be exactly the same. The growth standard provides no information as to how different the two centiles ought to be. Ironically, in growth terms the growth standard is no standard at all.

This is widely misunderstood. Changes in centile are often quantified in terms of the number of channel shifts occurring over a period of time, where a channel is the distance between two adjacent centile curves (usually about 0.6 of an SD score unit). Yet the centile curves are based on single measurements, and contain no information about the extent to which measurement centiles change over time. This form of classification is largely meaningless. Another problem is that channel shifts can be defined either in terms of the number of centile curves actually crossed (which depends on how close the child is to a centile curve initially), or else in units of the distance between neighbouring centile curves. The two definitions usually give different answers.

This distinction between size or status on the one hand, and growth on the other, needs to be reflected in the standards used to quantify them. There are essentially two ways this can be done, one due primarily to Tanner (1962) and the other to Emery *et al.* (1985).

A child's size at any age represents growth since conception, and it is a measure of how far he or she has travelled on the road to adult size. Tanner (1962) uses this analogy to refer to growth standards involving a single measurement as 'distance' standards, while standards based on growth measured over a period of time he terms 'velocity' standards. Fig. 4 shows median distance and velocity curves for girls' height (Tanner *et al.* 1966). The alternative curves are discussed in the next section.

There are important differences between distance and velocity standards, both in terms of their use and their construction. Velocity standards measure growth rather than status, and require two measurements instead of one, so they need to be constructed using longitudinal rather than cross-sectional data. Height velocity is calculated as $(V = (H_2 - H_1)/\Delta t)$, where H_1 and H_2 are the two measurements, and Δt is the time interval between them. Since each height measurement has its own measurement error, the variance of V is given by

$$\text{var}(V) = \sigma_v^2 + \frac{2\epsilon^2}{\Delta t}, \quad (6)$$

where ϵ is the measurement error and σ_v is the population standard deviation of true height velocity (Bairagi, 1986). This shows that the variability of velocity (and hence the spacing of the centile curves) depends on the time interval Δt between measurements. The velocity standard must therefore be tied to a particular time interval, to ensure that the two components of error are weighted appropriately. For height during childhood this is conventionally one year (Tanner *et al.* 1966), which has the added advantage of cancelling out any seasonal variation in height (Marshall, 1971). For length and weight during infancy a period of a year is too long, and intervals between 2 weeks and 3 months have been proposed (Emery *et al.* 1985; Bairagi, 1986; Healy *et al.* 1988*b*; Guo *et al.* 1991).

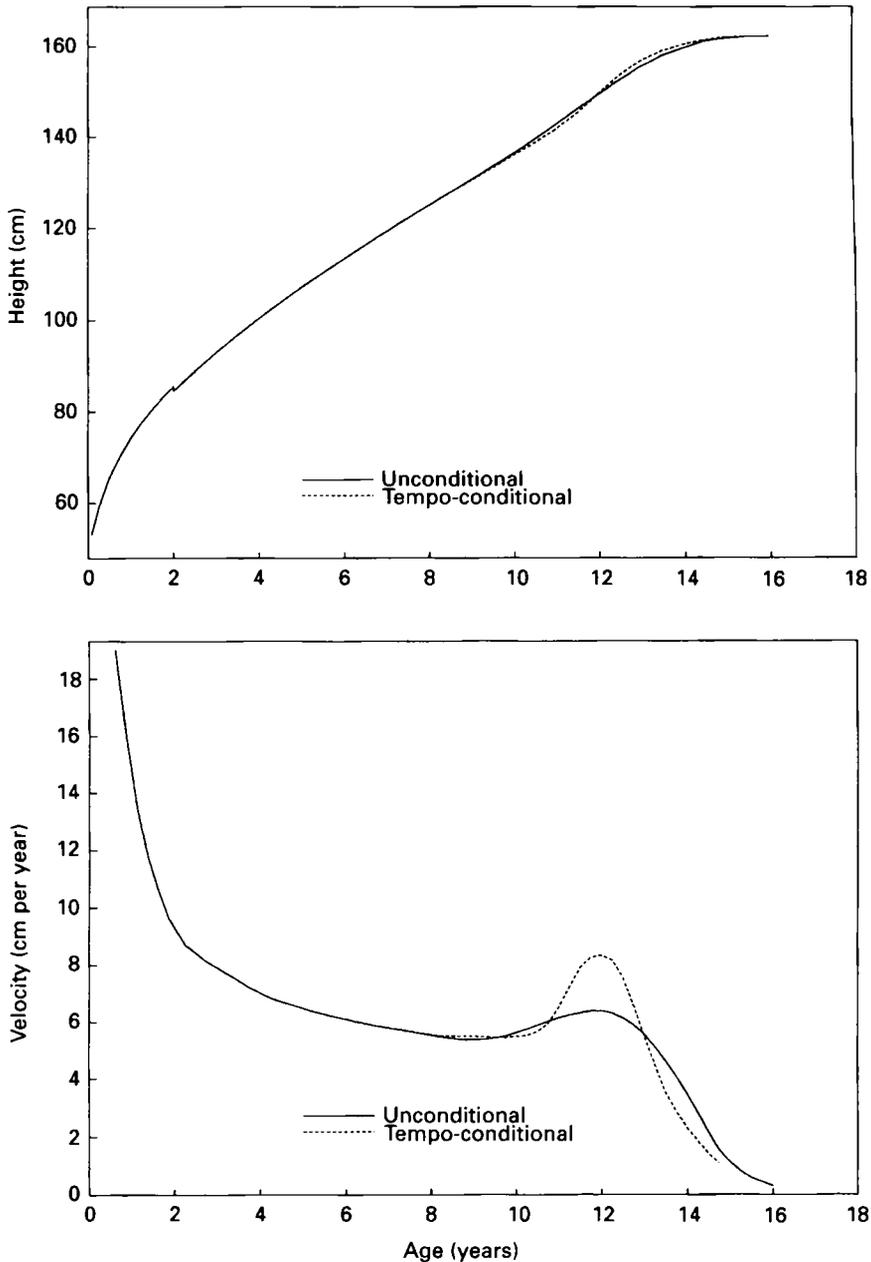


Fig. 4. Charts of girls' height distance and annual height velocity based on the Tanner Whitehouse 1966 standard. The charts show the difference between tempo-conditional (or clinical longitudinal) and unconditional standards.

The variability of height, i.e. the distance equivalent of equation (6), is given by $\text{var}[D] = \sigma_d^2 + \epsilon^2$, where σ_d is the standard deviation of true height and ϵ is the measurement error from (6). As with velocity, this includes two components of variation, but they have equal weightings. The signal to noise (S:N) ratio for distance and velocity can be defined as the

true variance divided by the measurement error variance. Taking a boy aged 8 years as an example, σ_d is 5.7 cm and σ_v is 0.8 cm/year (Tanner *et al.* 1966), while ϵ is 0.4 cm at best. So at this age the S:N ratio for height distance is about 200, while for height velocity over a period of a year it is 2, a hundred times smaller. So at this age velocity is very poor compared to distance for screening purposes. At 12 months the ratio of ratios is much smaller, at just over 3, so that distance and velocity are fairly similar. In general the ratio of the two S:N ratios is $2\sigma_d^2/\sigma_v^2$, which is independent of ϵ , the measurement error.

This discussion refers to normally growing children, and so is a statement about the relative specificity of distance and velocity, i.e. how accurately the two measures recognize healthy children. In terms of sensitivity – correctly identifying ill children – velocity is likely to be more effective than distance, in that it changes more rapidly during growth failure.

The velocity standard is one tool to quantify changes in measurement centile over time. The other is a distance standard marked with an extra set of centile-like curves (Emery *et al.* 1985; Cole, 1994), which indicate how much a child's centile can be expected to change over a given time period. Emery *et al.* (1985) developed the Sheffield Weight Chart on this principle; serial weights during infancy are plotted on a set of tramlines (strictly speaking they are not centiles), which are spaced so that there is a 5% chance of an infant's weight shifting the width of one tramline over a two-week period, or shifting two tramline widths over 8 weeks. In addition there is the usual set of weight distance centile curves. This form of chart is very powerful as it deals with both single and multiple measurements, and the significance of changes over time is shown graphically.

Weight for height

The primary purpose of growth standards is to adjust for age, but other forms of adjustment can be useful, e.g. weight adjusted for height (Hamill *et al.* 1977). The value of a weight-for-height standard is that it measures 'wasting' (Waterlow *et al.* 1977), while at the same time requiring no knowledge of the child's age. This is useful when the child's age is unknown, as is often the case in the developing world.

However, it has become clear that weight for height without reference to age can be misleading (Cole, 1985), and the BMI is preferred where age is known (Cole, 1979; Cronk & Roche, 1982; Rolland-Cachera *et al.* 1982). Standards for BMI versus age in children and adults from France and the USA have been published (Rolland-Cachera *et al.* 1984; Hammer *et al.* 1991; Must *et al.* 1991; Rolland-Cachera *et al.* 1991).

Chinn (Chinn & Morris, 1980; Chinn, 1992) has described two other approaches to weight for height that adjust for age, both applied to the age group 4–12 years. The first involves age-specific charts of weight for height, while the second uses the index $(\text{weight} - 9)/\text{height}^{3.7}$, with weight in kg and height in m. The index is less skew than the BMI, and has a constant cv.

MARGINAL AND CONDITIONAL STANDARDS

As described above, there are distance standards (based on single measurements) and there are velocity standards (based on pairs of measurements separated in time). Each type of standard may or may not use other available information.

Marginal or unconditional

A marginal or unconditional standard is constructed from a reference population where each individual contributes a single measurement, unadjusted for other information. This is by far the most common form of standard, and it expresses individual subjects in terms of a centile relative to the reference population on which the standard is based.

Marginal standards can be applied equally to distance and velocity – individuals from the reference population provide either a single measurement or a single velocity. Velocity standards in infancy tend to be unconditional (Guo *et al.* 1991), although in practice such infants are followed longitudinally, and often provide several velocity measurements. During puberty, velocity standards can be either unconditional or conditional (Tanner *et al.* 1966), as described in the next section.

Conditional

The alternative to the marginal standard is the conditional standard. It works on the principle that a child's measurement should be expressed conditional on, or adjusted for, another covariate in addition to age and sex. Conditional standards have been described for height and height velocity during puberty adjusted for tempo as defined in the next paragraph (Tanner *et al.* 1966), height conditional on mid-parent height (Tanner *et al.* 1970; Himes *et al.* 1981), birthweight adjusted for sibling birthweight (Tanner *et al.* 1972), height adjusted for height one year earlier (Healy, 1974; Cameron, 1980; Cole, 1994), and weight and height in early life conditional on previous measurements (Berkey *et al.* 1983).

The first of these, height and height velocity during puberty, are known as tempo-conditional or clinical longitudinal standards. The period of puberty involves a sudden acceleration in height followed by an equally sudden deceleration, and the time (or tempo) varies considerably from one child to another. Analysing such data cross-sectionally smooths out the details of the growth spurt, and the average curve (be it distance or velocity) is not the same shape as the individual curve for a typical child (Boas, 1930). Individuals plotted on the cross-sectional distance chart during puberty follow a steeper slope than the median curve, and on the cross-sectional velocity chart they have a higher and narrower peak of velocity, as shown in Fig. 4. Tanner (Tanner *et al.* 1966) modified the charts to take the variation in tempo into account. His conditional charts are effectively marginal distance and velocity charts for children of a particular age at peak velocity, so that their individual growth curves run parallel to the median curve. Children whose tempo differs follow similar paths, but shifted to earlier or later ages.

Unlike tempo-conditional standards, most other conditional standards are based on an underlying linear regression equation. The standard of height conditional on height a year earlier (Cameron, 1980) is an example. Consider two heights, H_1 and H_2 , separated in time. The linear regression of H_2 on H_1 is given by:

$$H_2 = a + bH_1 + \text{error}, \quad (7)$$

where b is the regression coefficient, a is the intercept, and the error term is obtained by minimizing the sum of squares of the residuals ($H_2 - a - bH_1$). The residual standard deviation is, say, s_c . The median of the standard of H_2 conditional on H_1 is a line given by the regression equation, while the other centiles are lines parallel to the median, the spacings being determined by the residual standard deviation s_c from the regression. For example, the 97th conditional centile is 1.88 s_c units above the median.

The intercept a in equation (7) is the same for all children, so a simpler form of residual for each child is given by $(H_2 - bH_1)$. This has a mean value of a and a standard deviation of s_c . It is a compact way of saying ' H_2 adjusted for H_1 ', and it is uncorrelated with H_1 by definition – if it were not, its variability could be reduced by altering b .

The value of b , the regression coefficient, is important. If it takes the value 1 the residual simplifies to $(H_2 - H_1)$, which is just the increment in height over time, corresponding to the velocity. Conversely, if b is very different from 1 the conditional standard and the velocity standard are quite different. Fig. 5 shows the value of b by age for boys and girls, based

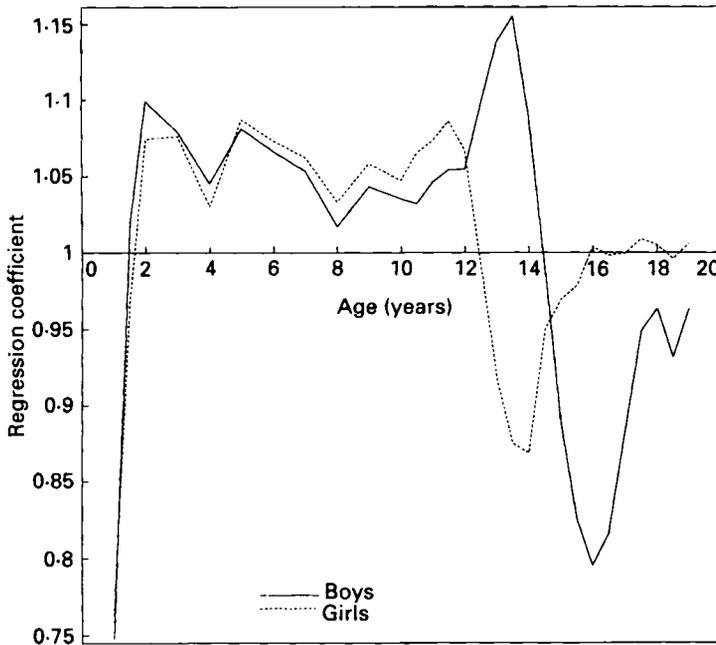


Fig. 5. The regression coefficient of height on height one year earlier during childhood, for boys and girls. The curves are based on data from the French Growth Study (Sempé *et al.* 1979; Cole, 1994).

on data from the French Growth Study (Sempé *et al.* 1979). The regression coefficient is below 1 in infancy and puberty, when centile crossing is relatively common (Cameron, 1980; Cole, 1994), while in mid-childhood it is just above 1.

A regression coefficient below 1 indicates appreciable regression to the mean. Individuals in the tail of the distribution on the first occasion are likely to be less extreme subsequently, so that those who start small grow faster, and vice versa. A conditional standard takes this association between distance and velocity into account, whereas a velocity standard does not.

The need for a conditional standard rather than a velocity standard during puberty has been emphasized repeatedly by Tanner (1962), and his clinical longitudinal and velocity standards achieve this. However, in infancy, the need for a conditional standard is less well recognized (Wright *et al.* 1994), even though catch-up growth is known to have a major effect on centile position (Smith *et al.* 1976).

Between infancy and puberty, b is slightly greater than 1. This reflects the fact that the centile curves are expanding, so that for a child to remain on, say, the 97th centile, he or she needs to grow faster than a median or 3rd centile child (Cole, 1994). In practice the association is small, so that an average child growing along the 3rd centile is on the 30th velocity centile (Bailey, 1994). There is less need for a conditional standard at this stage of childhood.

A conditional standard reduces the variability of the measurement. In the example above the standard deviation s_c of the residual ($H_2 - bH_1$) is given by $s_2\sqrt{1-r^2}$, where r is the correlation between H_1 and H_2 . Thus the conditional standard is useful when r is large. If it takes the value 0.8, the residual standard deviation is 0.6 s_2 , a 40% reduction compared to the unadjusted measurement. Height adjusted for mid-parent height, and birthweight adjusted for sibling birthweight, are two examples where conditioning leads to a more

precise estimate. Note though that the reduction in standard deviation depends on a high correlation – if the correlation falls from 0.8 to 0.6, the reduction is 20% rather than 40%. Conditional standards require a strong correlation between the measurement and the conditioning variable.

For conditional standards such as height on previous height, the comparison should be with the height velocity standard. In this case there is little advantage in terms of a reduced standard deviation – 5% or less for a correlation of 0.8. Here the benefit of a conditional standard lies in the reduced bias rather than the increased precision. Conditional predictions for individuals in the tails of the distribution are unbiased, whereas velocity standard predictions are often biased.

PURPOSE OF STANDARDS

The ways in which growth standards are used can be viewed in two broad categories: monitoring growth in individuals, and summarizing mean growth in groups. Within each category there are subcategories that apply to particular age groups and anthropometric measurements.

Education

In the developing world growth, particularly in infancy, is often compromised due to the effects of factors including poor hygiene, diet and infection. Providing mothers with information about their child's health is a powerful way of counteracting these influences as recognized by UNICEF (Grant, 1984). GOBI is UNICEF's acronymic name for a four-pronged strategy to influence child survival, of which the G stands for Growth Monitoring and Promotion, and the other components are Oral rehydration, Breast feeding and Immunization.

Growth monitoring is achieved through charts such as the 'Road to health' chart (Morley, 1973). This is a simplified weight chart, usually with just two curves representing the boys' median and the girls' 3rd or 5th centile from an international standard, and the aim is to keep the child growing along the road between the curves. Thus longitudinal measurements of weight in infancy are the primary focus, although other measurements such as length, arm circumference or head circumference are also used.

Screening

Anthropometry is widely used as a screening tool for disease in children and adults. Here the anthropometry acts as a proxy for disease – small size or poor growth does not of itself constitute disease, it is simply a marker for it. The process of screening involves identifying individuals below some prespecified cut-off, on the assumption that their chance of being ill is greater than for the population as a whole.

Screening is used in many different contexts, from refugee camp to school entry. In refugee camps the purpose is triage, to optimize the allocation of food aid by identifying children most in need. Weight adjusted for height is the usual criterion here, as age may be unknown. This process assumes that low weight for height predicts increased risk of mortality, which is true in the emergency situation (Toole & Malkki, 1992). In the longer term, weight for age and height for age perform better than weight for height (Chen *et al.* 1980; Bairagi *et al.* 1985; Pelletier, 1991), but the sensitivity and specificity of the screening process when applied in the community is low (Van Lerberghe, 1988).

Velocity screening during infancy is of value, both in the developing and the western world, for identifying failure to thrive. It is important to use the appropriate form of longitudinal standard, since this is the age group where centile shifts are often observed, and incorrectly interpreted, using a distance chart. Following the recommendations of the

WHO Working Group (WHO, 1986) there has been an upsurge of interest in the use of velocity measurements during infancy (Healy *et al.* 1988*b*; Guo *et al.* 1991; Piwoz *et al.* 1992). Previously there was not an appropriate standard, and there has always been the difficulty of providing data for several different time intervals.

For school entry, height screening is used to identify children with possible growth problems (Voss *et al.* 1992). The height centile used as the cut-off (usually the 3rd in the UK) defines the level of false positives (Healy, 1991), so that the specificity is set to 97%. In theory the sensitivity should be similarly high, with virtually all the abnormally growing children falling in the lower tail of the distribution. In practice, however, there will always be children whose growth problems have not yet surfaced, so that the sensitivity is difficult to quantify.

Screening of height velocity at school entry has been suggested, but the low signal–noise ratio of height velocity makes it unspecific. In addition, the measurement error of height at school entry is relatively high (Voss *et al.* 1990), which increases the false positive rate even more. One way forward is to insist on two successive height velocities below the cut-off (Healy *et al.* 1988*b*), as they are relatively uncorrelated with each other. For example, the chance of a normal child falling below the 10th velocity centile for two successive years is only 10% of 10%, or 1%. But even this does not approach the performance of screening based on a single measurement, and it also takes two years to reach an answer.

Clinical

The arena in which height velocity is of particular value is the specialist growth clinic. Children with a diagnosis of growth disorder need to be monitored to evaluate the effect of treatment, and height velocity is the direct measure of growth response. Measurement error in the growth clinic is lower than elsewhere, so there is the opportunity to detect a differential response.

In principle a conditional standard approach is more appropriate than a velocity standard, even here. For example, in idiopathic growth hormone deficiency treated with growth hormone, the growth response is known to be greater in children with a smaller initial height centile (Ranke & Gilbaud, 1990). The statistical advantage of the conditional standard is that it can be extended naturally to a full regression model with other covariates.

Group summary

Anthropometry is often used to summarize the growth status of groups, with the aim of identifying groups at particular risk of growth faltering. Use of a growth standard enables distinct age–sex ranges to be combined, which increases the ability to summarize the results. However, the standard needs to be appropriate for the data, or else the summary statistics may be misleading.

An extreme example of this would be the use of a growth standard to compare the efficacy of different treatments, say food supplementation or growth hormone therapy, in groups of children unmatched for age. If the standard were inappropriate for the data, it could show apparent differences in growth between the groups due to their differing ages which would, quite wrongly, be attributed to the treatment.

EFFECT OF AN INAPPROPRIATE STANDARD

Whatever type of growth standard is used, it is important to know the possible effects of using a standard which is inappropriate for the data. The mismatch between the standard and the data can be expressed in terms of the offset and the trend, as applied to a particular age range. The offset is defined as the mean SD score of the data, while the trend is the

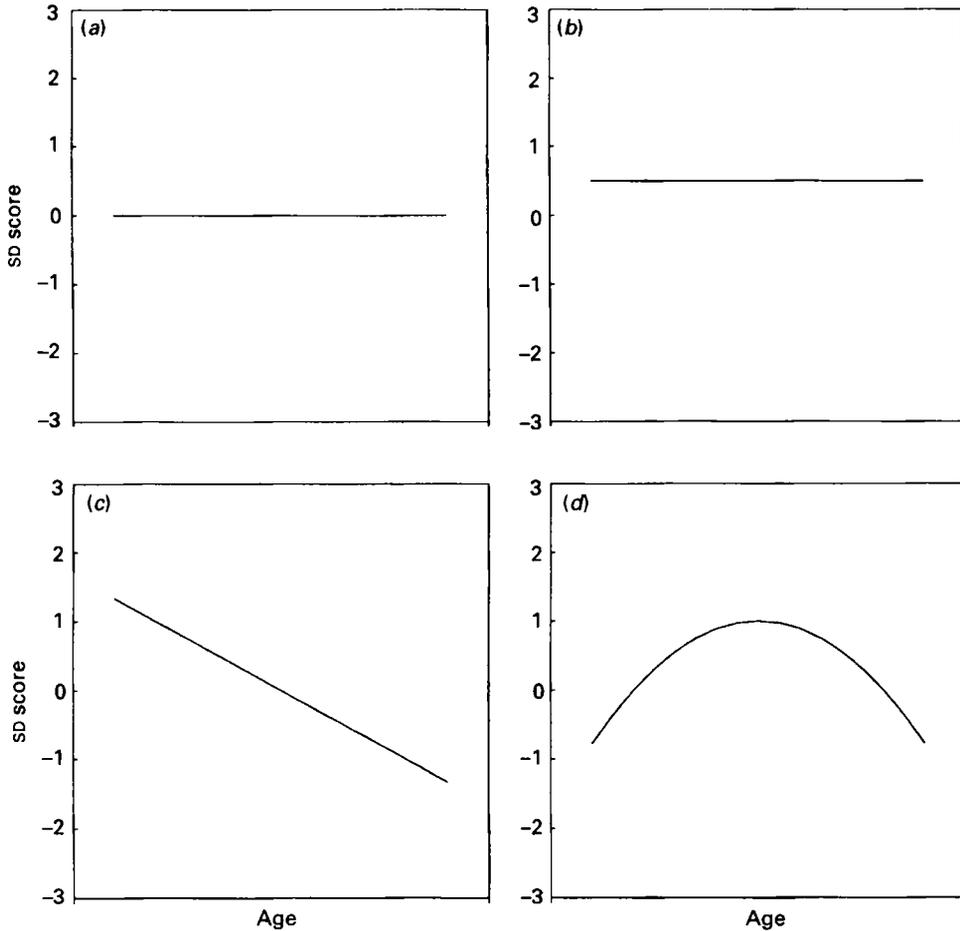


Fig. 6. Hypothetical data to show the effect of adjustment with an inappropriate growth standard. The SD score of the adjusted measurement is shown plotted against age: (a) is the ideal outcome, with no offset and no trend; (b) shows an offset but no trend; (c) is a constant negative trend and (d) is a varying trend.

regression coefficient of the SD scores of the data on age. If both the offset and the trend are near to zero, the standard and the data are well matched – this is the ideal (see Fig. 6a).

If the trend is small but the offset is appreciable (Fig. 6b), this suggests that the standard is inappropriate – the mean SD score of the data is non-zero. This might occur if, say, children from the North and South of England were compared. However, since the trend is zero the data are appropriately adjusted for age, and this may be quite sufficient for the particular purpose. The offset provides an age-independent summary of the data.

One particular case where a non-zero offset is important is screening for growth disorders. Applying a standard to children of the same reference population, the 3rd centile used as cut-off should provide a false positive rate of 3%. If there is a mismatch between standard and data (perhaps because the standard is out of date), the basis of the screening procedure is invalidated. Thus the proportion of British children falling below the 3rd centile of the Tanner standard is currently 1.3% (Voss *et al.* 1992), indicating that about 1.7% of the population is not being followed up simply because the standard is outdated.

A third possibility is that the trend of SD score with age is non-zero, so that the standard fails even to adjust for age (see Fig. 6*c*). In this case the value of the offset is meaningless, as it changes with age. This happens when a developed world standard such as the NCHS (Hamill *et al.* 1977) is applied to infant growth in the developing world. It converts the trend to increasing anthropometry during the first year to a trend of falling SD score (Rowland *et al.* 1977).

Another possibility involves the timing of puberty. If peak height velocity occurs earlier in the data than in the reference population, the mean SD score for height distance is higher at the reference age of peak height velocity, and subsequently lower for a short time. The mismatch in age at peak velocity generates a transient rise followed by a fall in mean distance SD score, as shown in Fig. 6*d*.

As another example, the pattern of growth during infancy in the Western world has changed significantly over the last 20 years (Whitehead & Paul, 1984). Modern infants assessed on the NCHS or Tanner–Whitehouse charts appear to rise across centiles for the first three months of life, but then to fall back again (Fig. 6*d*). This is a worrying pattern of growth for those who do not realize that it arises from the inadequacy of the standard.

Statistically speaking, these examples should not matter so long as the residual age trend is adjusted for. However, if the age trend is assumed to have been removed when it has not (as in Fig. 6*c*), it may generate a spurious correlation between the anthropometry and other age related variables. To be safe in such situations, age should always be tested for using regression. Furthermore, if infancy or puberty occur in the age group, a quadratic age trend may be required to compensate for patterns similar to Fig. 6*d*.

In clinical terms, an out-of-date growth standard causes problems (Voss *et al.* 1987; Chinn *et al.* 1989). Currently a new UK growth standard is being developed to replace the venerable Tanner–Whitehouse 1960s standard. The new standard will provide national height and weight distance charts, along with information about regional variation.

FUTURE DEVELOPMENTS USING SD SCORES

Display of SD scores

With recent statistical developments (Cole, 1988; Healy *et al.* 1988*a*) and the availability of desktop computers, it is now straightforward to convert anthropometry into SD scores. This is an important advance, because multiple anthropometry measures can be combined very easily – across measurements, across ages, across subjects – since SD scores are on a unified scale. Thus one can display, say, a series of height and weight measurements for a brother and sister, all on the same graph (Cole *et al.* 1989).

This form of presentation not only puts more information into the chart, it is also better suited to identifying growth faltering. Normal growth appears on the chart as a horizontal line, so a child whose centile position is slowly falling is more obvious on an SD score chart than on a conventional chart.

Specialist growth software is starting to appear which should simplify the use and construction of growth charts (Anon. 1993). Computers can easily calculate and print out a chart that is specific to an individual subject, for example, height adjusted for previous anthropometry (Berkey *et al.* 1983).

Indices based on SD scores

The conditional standard described by Berkey *et al.* (1983) provides a very powerful tool for the interpretation of longitudinal growth data. However, it makes the twin assumptions that height and log (weight) are normally distributed, which cannot be relied on (Cole,

1989*a, b*). Berkey's method is considerably simplified if, instead of using height and log (weight), it is based on the SD scores of height (Tanner *et al.* 1966; Cole, 1989*b*) and weight (Cole, 1989*a*).

Consider two anthropometry measurements, with SD scores Z_1 and Z_2 , with a correlation between them of r . The conditional standard of Z_2 on Z_1 is given by

$$Z_2 = a + bZ_1 + \text{error} \quad (8)$$

as for equation (7) above, where a and b are the intercept and coefficient from the linear regression equation. However, because Z_1 and Z_2 are SD scores, the equation can be simplified. If used with children appropriate to the underlying reference population, then the means of Z_1 and Z_2 are both 0, so that the intercept a is also 0 by definition. In addition, the standard deviations of Z_1 and Z_2 are both 1, and the regression coefficient b is equal to the correlation r by definition. So the equation simplifies to

$$Z_2 = rZ_1 + \text{error}, \quad (9)$$

the expected value of Z_2 is rZ_1 , and the expression $(Z_2 - rZ_1)$ is a measure of Z_2 conditional on Z_1 . The residual standard deviation of Z_2 , and hence the standard deviation of $(Z_2 - rZ_1)$, is given by $\sqrt{1 - r^2}$, since the standard deviation of Z_2 is 1. This allows a further simplification: if the index is divided by its standard deviation, it becomes an SD score in its own right. So the conditional index

$$Z_{2|1} = \left[\frac{Z_2 - rZ_1}{\sqrt{1 - r^2}} \right] \quad (10)$$

has a mean of 0 and a standard deviation of 1 – it is an SD score of Z_2 conditional on Z_1 .

This conditional SD score index is very flexible. The measurements providing the SD scores Z_1 and Z_2 are quite general; they could be two measurements on the same occasion (e.g. height and weight), or the same measurement on two different occasions, or even measurements from two different individuals (e.g. child and parent, or siblings). The first example corresponds to an index of weight for height, the second to an index of height conditional on previous height, and the third to parent-adjusted height or sibling-adjusted birthweight. In each case, the only information required is (a) the growth standard(s) to convert the measurements to SD scores, and (b) the correlation r between the two measurements. The required correlation will usually depend on the age and sex of the particular child and may well be available in the literature (Cameron, 1980; Berkey *et al.* 1983; Cole, 1986).

One area of assessment which could benefit from this form of index is growth velocity in infancy. Infants are usually measured at irregular time intervals, and the growth rate to expect between any two ages is generally not known. Current standards of weight and length velocity are geared to fixed periods of weeks or months (Emery *et al.* 1985; Guo *et al.* 1991). There is also the difficulty (discussed above) that velocity is related to distance in infancy, so that a conditional standard is preferable to a velocity standard.

The extra information required for the conditional index is the age-on-age correlation r . In general this correlation is higher the closer together in time the two measurements are, and it increases with increasing age (Berkey *et al.* 1983). It is also higher for length and height than for weight. It would be straightforward to construct a mathematical model for r , using existing longitudinal data, which would then predict the appropriate correlation for any pair of ages. Using this model, an infant followed up at a growth clinic could be assessed using the conditional SD score index based on equation (10).

A second use of the index is to measure weight for height as an alternative to the BMI. It happens that for much of childhood after infancy until puberty, the correlation between

age-adjusted height and weight is close to 0.7 (Cole, 1986). Substituting this correlation into equation (10) and rounding gives the index $1.4 Z_2 - Z_1$, where Z_2 is weight and Z_1 is height. It shows that children on the median for weight and height (i.e. $Z_1 = Z_2 = 0$) are also on the median for weight for height. The same is true (approximately) for children on the 97th height centile and 90th weight centile, or the 3rd height centile and 10th weight centile. Perhaps surprisingly, children whose height and weight centiles are the same are generally not median weight for height unless they are also median weight and median height.

The conditional index is designed to be distributed as an SD score, with a mean of 0 and an SD of 1. However, it assumes that the parent SD scores Z_1 and Z_2 are distributed as bivariate normal with mean 0 and SD 1, and that the value of r is appropriate for age and sex. If these assumptions are not met, one or more of the following will occur: the index will have a non-zero mean, a non-unity SD, a non-normal distribution or a non-zero correlation with Z_1 . The error in r will be of greater significance the closer r is to 1.

A relevant consideration here is measurement error – the bigger it is, the smaller the correlation. So modelling the correlation in, say, a research study, and then applying it to routinely collected data, may lead to poorly specified SD scores. In such cases it may be possible to ‘shrink’ the correlation to take into account the increased measurement error.

GROWTH STANDARD CONSTRUCTION PRINCIPLES

Constructing a growth standard involves several distinct operations – identifying the reference population, making the measurements, and then calculating the required centiles. Waterlow *et al.* (1977) have provided a useful set of criteria for the separate stages of the process. In brief they require the reference population to be well nourished, the sampling procedure to be clearly defined and reproducible, the sample to be cross-sectional and of adequate size, the measurements to include all those that are relevant and to be of good quality, and finally the data and the smoothing procedures to be available. None of these requirements is controversial, except in detail, and yet they rule out most existing standards (Johnston & Ouyang, 1991). The exceptions are mainly standards derived from national stratified random samples, such as those of Cuba (Jordan *et al.* 1975) and Holland (Roede & Van Wieringen, 1985). The British Tanner–Whitehouse and the American NCHS standards both fail because the reference data include a longitudinal component.

The issues of choice of population, sampling procedure and measurement technique are not pursued further here, as excellent accounts are available elsewhere (Cameron, 1986; Goldstein, 1986). The rest of the review concentrates on the statistical procedures available for converting a set of anthropometric data into a centile chart. This involves two distinct concerns – the shape of each centile curve, and the spacing between curves.

GROWTH CURVES

Smoothing

Modelling of the human growth curve can be applied to individuals or to group means. Either way, the shape of the curve is complex. In the past the most effective way to model it was to draw it by hand – this had the benefit of flexibility, but at a cost of subjectivity. With the advent of computers, the drawing hand has been replaced by a cubic spline or kernel smoothing procedure (Stützel *et al.* 1980; Cole & Green, 1992), which has reduced but not eliminated the element of subjectivity. It is important to remember that smoothing

involves a trade-off between smoothness and goodness of fit, and there is no single correct curve. Splines and kernels are essentially moving averages of the data (Guo *et al.* 1990a).

One minor disadvantage of free-form curves, be they drawn by hand or computer, is that they are cumbersome to describe mathematically – they have to be specified as a table of values. Another weakness is that in situations where the form of the curve is known, for example asymptotic growth in height, the curve cannot easily be constrained appropriately. An alternative is to express the growth curve as a simple mathematical function. Many different types have been suggested over the years, falling into two distinct classes: those applicable to the early part of life, when growth velocity is falling steadily, and those covering the whole of childhood including the pubertal growth spurt.

Infancy models

The simplest function for any period of childhood is the polynomial in age; it is easy to fit, and the degree of complexity can be extended indefinitely by increasing the order of the polynomial. However, polynomials have bad habits – they are restricted in the shape of curve they can model, their behaviour at the extremes of the data is unpredictable, and the addition of higher-order terms does not guarantee a suitable curve. The growth process is essentially smooth and gradual, and exponential functions are generally more suitable than polynomials for modelling it.

Two early models of infant growth were those of Jentsch & Bayley (1937):

$$y = a + bt - \exp(c + dt) + \text{error}$$

and Count (1943):

$$y = a + bt + c \log(t) + \text{error},$$

where y is the measurement (height or weight), t is age, $\log(t)$ is the natural logarithm of age, and a , b , c and d are parameters to be estimated. Both models combine a linear and a decaying age term, which means that their first derivative, the velocity, falls steeply initially and then flattens off. The Jentsch–Bayley model, with 4 fitted parameters, provides a better fit than the 3-parameter Count model (Berkey, 1982).

Several variants of the Count model have been proposed, involving different power transformations of age. Berkey & Reed (1987) suggest the Reed family of models, which add a series of inverse age terms:

$$y = a + bt + c \log(t) + d/t + e/t^2 + \dots + \text{error},$$

while Guo *et al.* (1990b), modelling weight in infancy, replace linear age with the square root of age:

$$y = a + b\sqrt{t} + c \log(t) + \text{error}.$$

All three models are special cases of the family of fractional polynomials described by Royston & Altman (1994). A fractional polynomial is a regression function where the terms are powers of age, but the choice of powers is widened to include negative as well as positive integers, and in addition 0.5 (square root), 0 (natural log) and -0.5 (inverse square root). The model combines power terms as necessary, so that the Guo model above has the two powers 0.5 and 0. Even with only two power terms, the form of the fitted curve is very flexible, and in many cases a single term is sufficient. The fractional polynomial family is likely to have very wide applicability in growth curve analysis.

The ICP model (Karlberg, 1987) quantifies the human growth curve by splitting it into three sections, Infancy, Childhood and Puberty (hence ICP). For the infancy component Karlberg uses the 3-parameter model

$$y = a + b[1 - \exp(-ct)] + \text{error}.$$

The Kouchi model (Kouchi *et al.* 1985) is another 3-parameter model of the form

$$y = a + bt^c + \text{error}.$$

Simondon *et al.* (1992) have compared the Count, Reed, ICP and Kouchi models for weight during infancy, and find that the ICP is the best-fitting 3-parameter model, while the 4-parameter Reed model is the best overall. However, they warn that their conclusions probably do not apply generally.

There are many variables that affect the fit of a model – the measurement, the age range and the children's growth pattern – so different models are likely to be optimal in different situations. However, a general conclusion is that during early childhood about 4 parameters are needed to model growth adequately.

Models that contain logarithmic or inverse age terms cannot handle zero or negative ages. A constant interval, e.g. 1 month or 9 months, can be added to each age to get round this. However, the choice of constant affects the fit of the model, so it is possible to improve the fit by optimizing the constant. In this case the model's number of parameters is effectively increased by one. Other models by contrast, for example the Jenss–Bayley, are invariant to changes in age scale and origin.

All childhood models

Moving from infancy to all childhood, the complexity of the model required increases substantially. The first derivative of the model should match the pattern of height velocity, which falls through childhood until puberty and then peaks briefly (Fig. 4). The earliest parsimonious model to achieve this was the 5-parameter curve of Preece & Baines (1978):

$$y = A - \left\{ \frac{B}{\exp\left[\frac{t-E}{D_1}\right] + \exp\left[\frac{t-E}{D_2}\right]} \right\} + \text{error},$$

which has been widely applied since. The parameter A is adult height, B is another height, E is an age offset, and D_1 and D_2 are rate constants. However, the fit of the model is poor below about 5 years, and it gives a biased estimate of the age at peak height velocity (Jolicoeur *et al.* 1988; Guo *et al.* 1992).

Another early height model was the triple logistic (Bock & Thissen, 1976), requiring 9 parameters:

$$y = \frac{A_1}{1 + \exp\left[-\frac{t-E_1}{D_1}\right]} + \frac{A_2}{1 + \exp\left[-\frac{t-E_2}{D_2}\right]} + \frac{A_3}{1 + \exp\left[-\frac{t-E_3}{D_3}\right]} + \text{error}.$$

This has not proved as popular as the Preece–Baines model. More recently, Shohoji & Sasaki (1987) have proposed a general 6-parameter function of the form

$$y = Af(t) + [1 - f(t)]c(t) + \text{error},$$

where $f(t)$ is the Gompertz function $f = \exp(-\exp(a - bt))$, $c(t)$ is the Count function (see above), and A is adult height. The model is thus a weighted combination of infant growth $c(t)$ and adult height A , where the weighting function is effectively zero until about age 10, and then rises steeply until adulthood.

In practice other infancy models, which are known to fit better, can be used here instead of the Count model (Jolicoeur *et al.* 1992; Simondon *et al.* 1992). The Jenss–Bayley model (unpublished observations) or an extended Count model (Jolicoeur *et al.* 1992) provide a fit over childhood almost matching that of the JPPS model described below.

One slight disadvantage of the Shohoji–Sasaki model is that its value peaks at about age 20, and then declines again. This is in contrast to the other childhood models, which tend to an asymptote representing adult height.

The ICP model of Karlberg (1987) has three separate components representing infancy, childhood and puberty. The childhood component is linear in age, while the pubertal component is logistic. A weakness of the method is that each component of the model is fitted separately, whereas other models are fitted over the whole age range. However, there are strong similarities between the ICP model and the Shohoji–Sasaki model modified to use Jeness–Bayley during infancy. The latter model, with 7 parameters, is as follows:

$$y = [1 - f(t)][a + bt - \exp(c + dt)] + f(t)A + \text{error},$$

where $f(t)$ is the Gompertz function, similar to the logistic. Since $f(t)$ is very small for $t < 10$ years, and $\exp(c + dt)$ is very small for $t > 5$ years, the equation can be rewritten as:

$$y = a - \exp(c + dt) + bt + f(t)[A - a - bt] + \text{error},$$

which is a combination of infancy, childhood and puberty components. Thus the ICP model can be fitted explicitly by viewing it as a combined Jeness–Bayley and Shohoji–Sasaki model.

A completely new form of 7-parameter model was described by Jolicoeur *et al.* (1988):

$$y = A \left\{ 1 - \frac{1}{1 + \left(\frac{t}{D_1}\right)^{c_1} + \left(\frac{t}{D_2}\right)^{c_2} + \left(\frac{t}{D_3}\right)^{c_3}} \right\} + \text{error}$$

where age t is measured from conception rather than birth. The JPPS model performs considerably better than the Preece–Baines, Shohoji–Sasaki and triple logistic models, but even so the authors have recently suggested ways of improving it (Jolicoeur *et al.* 1992). In contrast, Guo *et al.* (1992) recommend the triple logistic model as an alternative to kernel regression, although they do not consider the JPPS model. Figure 7 illustrates the fit of the JPPS model to the 37 tabulated values of the Tanner–Whitehouse longitudinal girls' height standard from 1 month to 17 years. The residual standard deviation is 0.6 cm.

All the childhood models cope well with the adolescent growth spurt, but they are less good with the mid-growth spurt, a phenomenon seen in some children between infancy and puberty. The only models that allow for a convex velocity curve during this phase of growth are the triple logistic and the JPPS, but even they do not provide a particularly good fit.

The conclusion is that the human height growth curve, to be modelled appropriately, would require at least 8 degrees of freedom.

CENTILE SPECIFICATION

Growth charts are like fishing nets. The centile curves are constrained by links in both horizontal and vertical directions, the former representing consistency with changing age, and the latter a well behaved distribution within age. The first of these constraints is explicit, in that the curves are smoothed across ages. The second constraint, that the frequency distribution of the measurement at each age should be consistent in some sense, and hence that the centile curves should be consistently spaced relative to their neighbours, is less generally accepted. Workers in the field of centile curve estimation are split on the issue – either the distribution at each age can be assumed to be of simple form and hence can be estimated, or else no distributional assumption at all can be made. Cole & Green

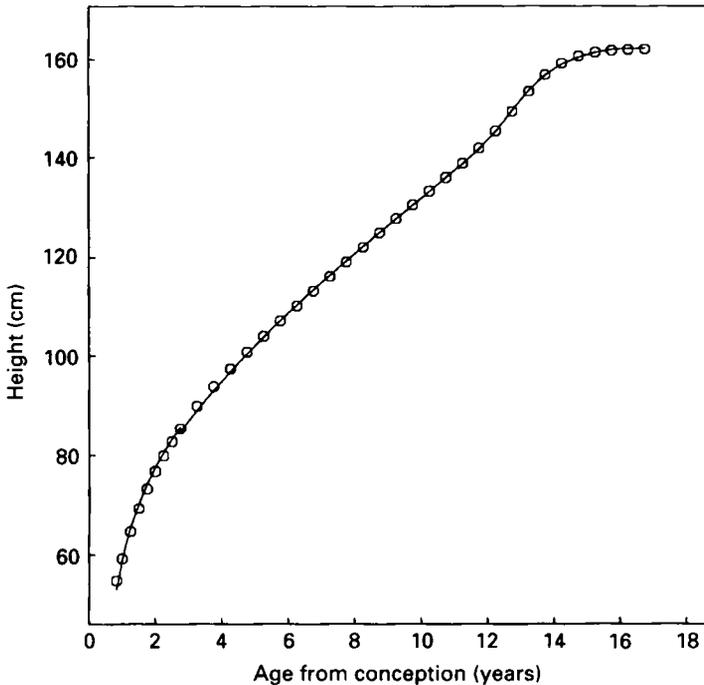


Fig. 7. The 7-parameter JPPS model (Jolicoeur *et al.* 1988) fitted to the Tanner-Whitehouse girls' tempo-conditional height median. There are 37 points, ranging in age from 1 month post term to 17 years, and the residual standard deviation is 0.6 cm.

(1992) use the term 'commonality' to describe centile curves being linked in position to their neighbours.

Normal distribution

The simplest example of commonality arises when the centiles are normally distributed. This is the approach used for the majority of height standards. The advantage of the normal distribution is that it defines the centiles in terms of the mean and standard deviation (SD) of the distribution:

$$\text{measurement centile} = \text{mean} + \text{SD} \times Z$$

where Z is the normal equivalent deviate (NED) for the required centile. Thus for the 90th centile the corresponding value for Z is 1.282, so that the 90th centile is 1.282 times the standard deviation above the mean. In terms of centile construction, the major advantage of assuming a normal distribution is that the standard errors of the estimated centiles are greatly reduced, particularly in the tails of the distribution (Healy, 1974).

To obtain smooth centile curves, the mean and SD of the distribution are calculated for narrow age-sex groups, adjusted for the grouping (Healy, 1962) and then smoothed (Tanner *et al.* 1966). Age-sex-specific values for the mean and SD are then substituted in equation (11) to obtain the required curves, which are themselves smooth by definition. Techniques for doing the smoothing are discussed above.

It has been known for over a century (Bowditch, 1891) that height is not normally distributed throughout childhood. There is a time during puberty when the distribution becomes skew to the right (Cole, 1989*b*). To compensate for this in the Dutch national

height standard, Roede & Van Wieringen (1985) smoothed the mean and SD, and then adjusted the centiles during puberty to fit the empirical centiles more closely.

More recent developments in this area have involved modelling the SD as it changes with age. Bland *et al.* (1990), deriving birth weight centiles for gestational age in very premature babies, observed that although the SD of birth weight increased sharply with gestation, the coefficient of variation ($CV = SD/mean$) was essentially constant. They proposed working with the natural logarithm of birth weight, as the SD of $\log(\text{birth weight})$ is equal to the CV of birth weight. This provides an efficient estimate of the CV, based on the whole dataset.

Another application of the same approach is described by Chinn (1992). At primary school age prior to puberty the SD of height is linearly related to the mean, so that data over a wide age range can be combined in the fitting process to give more precise estimates of the age specific means and SDs.

Altman (1993) proposes an elegant solution to the general problem of modelling the SD as it changes with age, based on the idea of absolute residuals. He first models the mean of the distribution (Royston & Altman, 1994), and then subtracts the mean from each point to obtain the absolute residuals, ignoring the sign. The mean of the absolute residuals at any one age is proportional to the SD, so the residuals can be modelled as if they were the SD. This leads to two parsimonious functions, one defining the mean and the other the SD (Royston & Altman, 1994).

Transformed normal distribution

If the assumption of a normal distribution is invalid, the extreme centiles will be biased. However, the evidence for or against normality lies mainly in the tails of the distribution, and as Healy (1988) has pointed out, these are the parts of the distribution with the least data. In practice, if the form of the distribution is to be estimated from the data, there is only a limited amount of information available on which to base it. This suggests that a conservative approach, using some form of modification of the normal distribution, is a logical first step. It is also important that whatever distribution is used it must allow measurements to be converted into centiles and vice versa.

The important distributional feature of anthropometry such as weight is its skewness – the right tail of the distribution is longer than the left. This means that centiles above the median on the chart are more widely spaced than those below the median (see Figs 1 and 2). A common solution is to work with $\log(\text{weight})$ (Berkey *et al.* 1983), relying on the twin assumptions that the degree of skewness is the same at all ages in both sexes and that a log transformation removes it. However, neither assumption is true. A more rational solution is to use a general transform, which allows the degree of skewness to vary with age and sex in the same way as the mean and the CV do.

Assume that the relevant anthropometry measure, weight say, has a skew distribution, and that with a suitable power transformation (Box & Cox, 1964) the skewness can be removed. The Box-Cox family of transformations deals with a wide variety of skew distributions, and its degree of transformation is specified by the value of the power λ . Certain values of λ have familiar meanings; when $\lambda = 1$, the transform has no effect – it leaves the measurement unchanged; the value $\lambda = 0$ corresponds to the natural log transform $\log_e(\text{weight})$, and $\lambda = -1$ is the inverse or reciprocal transform $1/\text{weight}$. The effect of the latter two transforms is to give a new measure, transformed weight, which is less skew and hence nearer to a normal distribution than the original.

In practice, the value of λ required to remove the skewness is unlikely to be a whole number, and in addition it changes with age. This is the basis of Cole's LMS method (Cole, 1988; Cole & Green, 1992). The LMS method calculates values of λ at different ages, and summarizes them with a smooth curve called the L curve (L for lambda). The M curve (the

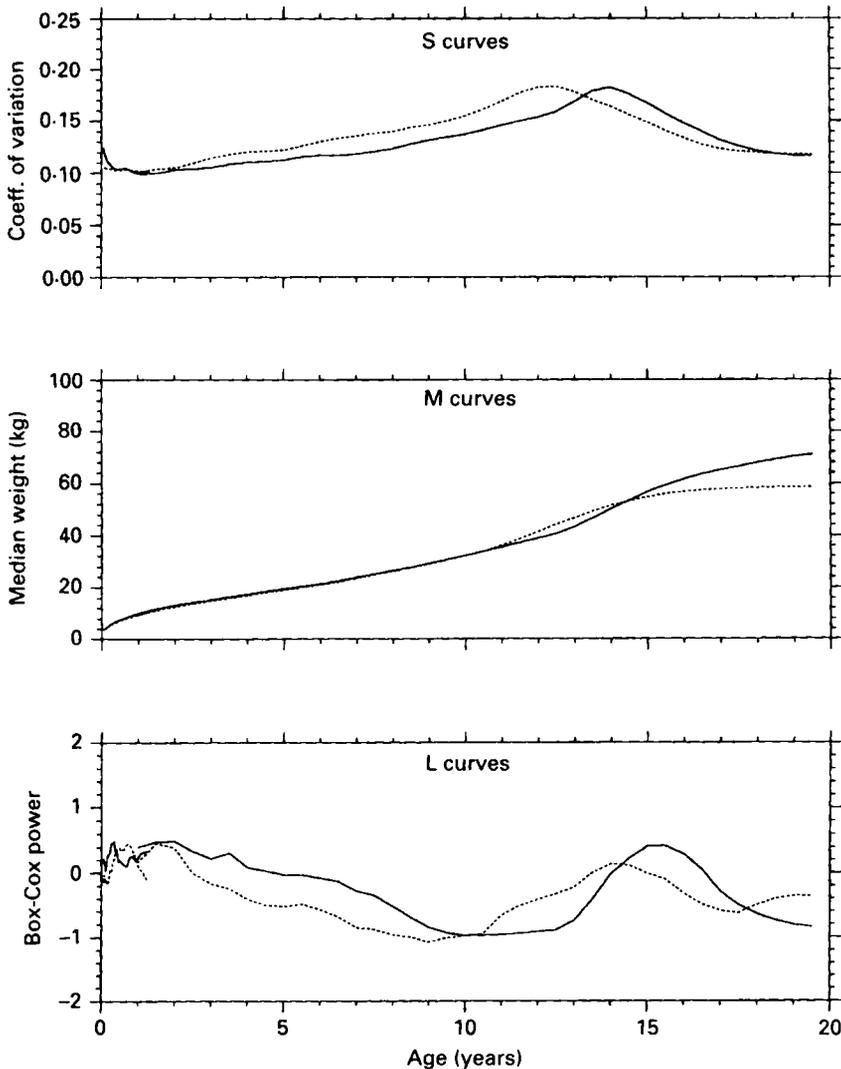


Fig. 8. The Dutch weight standard characterized by LMS curves, plotted against age by sex (boys ----, girls - - - -). The L curves indicate skewness, as measured by Box-Cox power transform, the M curves are median weight, and the S curves show the coefficient of variation.

median plotted against age; M for mu) and the S curve (the coefficient of variation by age; S for sigma) are derived simultaneously, and together they define centile curves for the growth chart. Each centile curve is defined as follows:

$$\text{measurement centile for age and sex} = M(1 + LSZ)^{1/L}, \quad (12)$$

where L , M and S represent values read from the appropriate curves for the appropriate sex at a series of ages, and Z is the normal equivalent deviate corresponding to the centile. If the three curves are smooth, the centile curves will be as well.

However, removing the skewness does not guarantee a normal distribution. The distribution will certainly be nearer to normal, with the mean and median closer together on the transformed scale than on the original scale, but there is no certainty that the higher

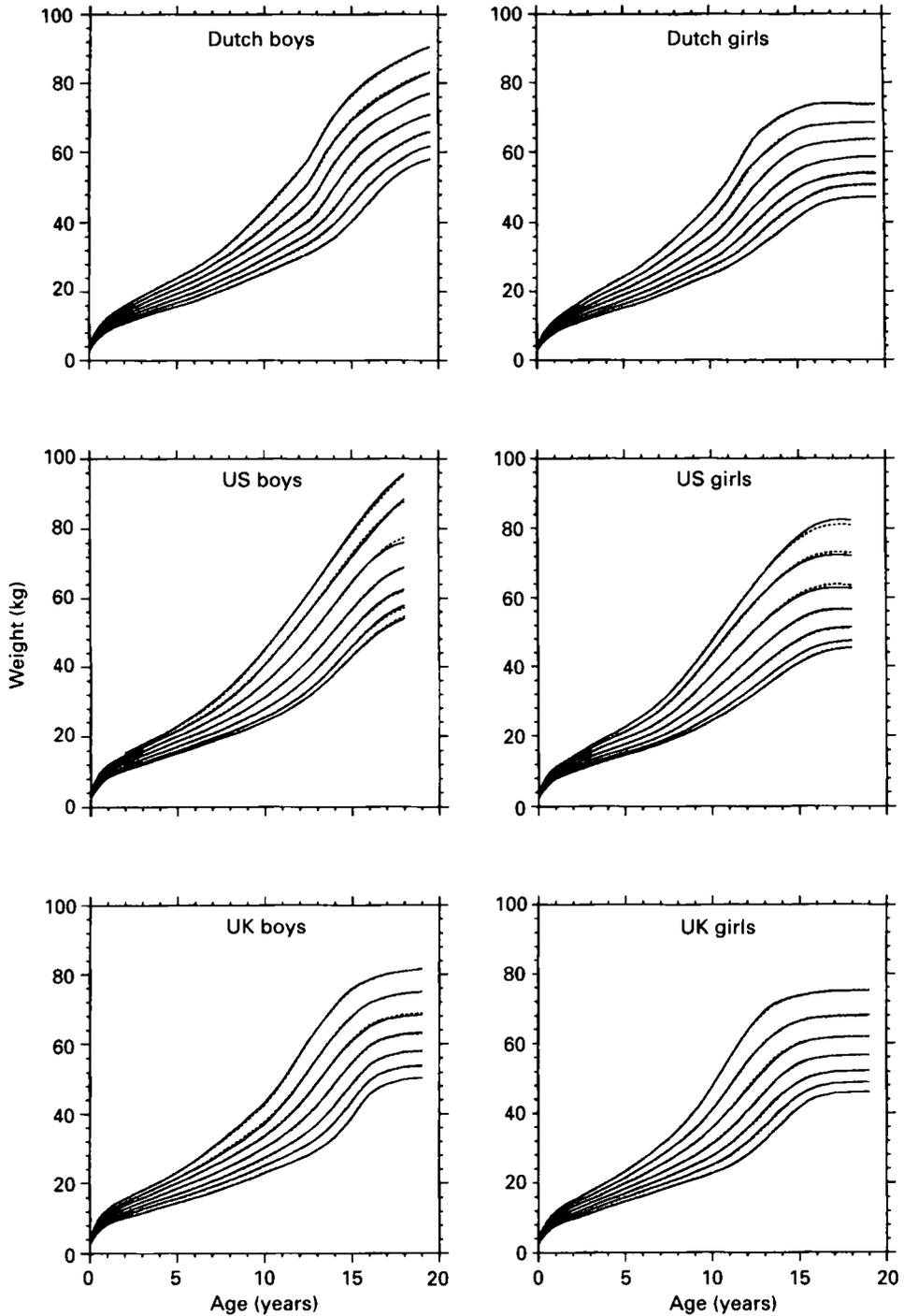


Fig. 9. A comparison of published centiles (solid lines) and LMS based centiles (dotted lines) for UK, US and Dutch national weight standards by sex (Tanner *et al.* 1966; Hamill *et al.* 1977; Roede & Van Wieringen, 1985).

moments of the distribution, notably the kurtosis, will also be normal. This is the key assumption of the LMS method, that once the skewness has been removed, what remains is acceptably close to a normal distribution.

Fig. 3 shows the effect of skewness on the centiles of an otherwise normal distribution, as the value of λ varies. The centiles are expressed as percentages of the median, which makes the median constant at 100%. It can be seen that values of λ in the range from -1 to $+2$ cover a wide spectrum of distributional shapes.

A benefit of the LMS method is that existing growth standards can be closely approximated by LMS curves fitted to the published centiles (Cole, 1989a). Fig. 8 shows calculated L, M and S curves for the Dutch weight standard by sex (Roede & Van Wieringen, 1985). They were originally published by Cole (1989a), but have been recalculated (Davies *et al.* 1993). They show that the CV of weight rises steadily until puberty and then falls again (S curves); that the power transform required to remove the skewness is, for much of childhood, more extreme than logarithmic (i.e. the L curves are less than zero); and that the skewness alters transiently during the pubertal growth spurt.

Fig. 9 shows the level of agreement between the published and LMS derived centiles, for three national weight standards (Tanner *et al.* 1966; Hamill *et al.* 1977; Roede & Van Wieringen, 1985). The discrepancies are generally small, with a standard deviation on the SD score scale of less than 0.03 units. As centile curves are typically spaced about 0.6 units apart, this level of discrepancy is in practical terms trivially small.

There have been other proposals to transform anthropometry to a known distributional form. Healy (1992) compared the performance of the shifted lognormal transformation $y = \log(x + c)$ with the Box-Cox power transformation $y = x^\lambda$ for summarizing published centiles. The advantages of the lognormal are that, like the Box-Cox, it includes the normal and the lognormal as special cases, while unlike the Box-Cox it can cope with negative values of x , which can arise with velocity measurements. It would be straightforward to develop a system analogous to the LMS method based on the lognormal, where the three age-specific curves would correspond to the mean and SD of the distribution, along with the value of the offset c . This would probably perform as well as the LMS method, although the interpretation of the three curves might be less obvious.

A more complex alternative is to use four parameters to summarize the distribution at each age. During the development of the American NCHS standard (Hamill *et al.* 1977), efforts were made to fit the distribution of weight at each age to a four-parameter Pearson distribution. In practice this did not work well, and the authors ultimately settled for a different form of analysis (see later). More recently Thompson & Theron (1990) used the family of 4-parameter Johnson curves to fit centiles to the distribution of birth weight by gestational age.

One reason why the NCHS initiative failed may have been that there was insufficient information in the sample adequately to specify a 4-parameter distribution at each age. The sampling error of λ in the LMS method is substantial (Cole, 1990), and this is only the third parameter (after the mean and CV). It implies that a fourth parameter, to be sufficiently specified, would require considerably larger sample sizes than are currently available.

Modelling skewness is an obvious next step after modelling the mean and variance. The weight distribution is widely recognized as being non-normal, so some form of extra flexibility has to be built into the distribution. Whether the flexibility is provided by Box-Cox, by Pearson or by Johnson is of less importance than the principle of fitting a family of distributions.

Constrained non-parametric

If explicit distributional assumptions are felt to be unacceptable, what other possibilities are there? The centile curves have to be estimated and smoothed independently of their neighbours.

There is in practice a compromise, where information about neighbouring centiles is brought into the estimation procedure, but without making distributional assumptions. Healy *et al.* (1988*a*) have described an ingenious two-stage method, where the rough centile curves are first obtained by a modification of the scatterplot smoothing technique of Cleveland (1979). The smoothing stage then exploits the fishing net quality of centile charts by smoothing the centiles in two directions, both across ages and within age across centiles. This is achieved by a constrained set of polynomial equations in age, where the coefficients are themselves polynomial equations in Z , the normal equivalent deviate for each centile.

In principle this method ought to be more flexible than the 3- or 4-parameter distribution families discussed in the previous section. The degree of the equation in Z corresponds to the shape of the distribution, so that a linear equation fits a normal distribution, a quadratic is approximately equivalent to a skew-normal distribution, and a cubic adjusts for kurtosis as well as skewness. However, Healy (1992) has shown that, in practice, a quadratic in Z does not adjust for skewness as effectively as a power or a log transformation. Furthermore, as discussed above, there is usually insufficient information in the sample to fit a fourth parameter to the distribution.

Another potential weakness of the method is the use of polynomials for smoothing: over wide age ranges they can fit poorly. However, this has been addressed by extending the method to use grafted polynomials (Pan *et al.* 1990).

Unconstrained non-parametric

The most extreme policy for fitting centiles is that information about neighbouring centiles should be excluded from the estimation procedure, so that the positioning of centiles is unconstrained.

Rossiter (1991) has proposed a variant on the idea of fitting a distribution at each age, but unlike those of previous sections, her method involves estimating a bivariate distribution by kernel methods, so that its shape can be quite general. Her kidney lengths example is unusual in that birth weight rather than age is the covariate.

The technique used by Hamill *et al.* (1977) to fit centiles to the American NCHS standard was simpler than Healy's method. The empirical centiles in each age group were smoothed across ages using a cubic regression spline, and each centile curve was estimated independently of its neighbours. As the sample sizes were not very large, this led to inconsistent centile spacings (Fig. 9).

A more sophisticated version of this approach was suggested by Jones (Jones, 1988; Jones & Hall, 1990). Non-parametric regression is used to estimate the centiles, and the degree of kernel smoothing applied to each centile curve is the same, which should make them moderately similar in shape. But again there is nothing in principle to stop adjacent centile curves touching or crossing.

A problem with non-parametrically estimated centile curves is the absence of a simple rule relating the centile values at any particular age with the corresponding Z values. Thus individual child measurements cannot easily be converted into SD scores. The NCHS recognized this (Hamill *et al.* 1977), and approximated the centiles by assuming them to be normally distributed, but with different standard deviations above and below the median. This demonstrates that centile curves, to be useful, need to have a distributional form forced on them at some stage in the estimation process.

Measurement error

The issue of measurement error is of considerable importance when constructing growth standards. It was shown above that velocity standards are more sensitive to this than distance standards, but the problem is quite general. If standards are used in a situation where the measurement error is appreciably larger or smaller than for the reference sample, misclassification will occur. This is the justification for measuring height velocity over a fixed period of a year.

It has recently become a particular problem with velocity standards during infancy, where the data are smoothed with a suitable growth curve, and the centiles constructed from the fitted curves (Roche *et al.* 1989). This process eliminates both measurement error and short term variation from the data, and the net effect is that the estimated centiles are too close together and growth deviations are exaggerated (Piwoz *et al.* 1992). It is possible to compensate for this by expanding the centiles relative to the median, using the multiplier $\sqrt{1 + (\epsilon/\sigma)^2}$ where ϵ is the measurement error and σ is the population standard deviation. The appendix of Piwoz *et al.* (1992) gives details.

CONCLUSIONS

USE OF STANDARDS

Centile charts will no doubt continue to be valuable graphic aids for the monitoring of child growth. Marginal distance standards are the most widely used form, where their simplicity is attractive. However, with recent advances in statistics and computing, more complex conditional standards are likely to become available, exploiting the use of SD scores. The ultimate aim must be an expert system for the evaluation of growth, which could be used for example to screen the large quantities of longitudinal preschool anthropometry now available in regional health authority computers throughout the western world. This would also have a spin-off in the developing world, where the potential benefits should be even greater.

CONSTRUCTION OF STANDARDS

There have been several recent advances in the construction of centile standards from reference data. To be of practical value, centiles need to be connected explicitly to the underlying Z value, which requires a distributional form for the centiles. The two approaches satisfying this requirement which have received the greatest attention are those of Healy (Healy *et al.* 1988a; Pan *et al.* 1990, 1992) and Cole (Cole, 1988; Cole & Green, 1992). Healy's method has been widely used for preschool data, helped by the availability of suitable software. It is also parsimonious in the way it defines the centile curves, requiring just the polynomial coefficients. However, it is restricted by its reliance on polynomials, so that more complex curve shapes, covering all of childhood for example, are fitted less well.

Cole's LMS method is the opposite: its use of cubic splines allows complex growth curves to be modelled, but they cannot be expressed in closed form. Also the program required to fit the cubic splines is not generally available, although a simpler method has been described (Cole, 1990). Both methods allow centiles to be converted into SD scores and vice versa, but the LMS method is simpler in this respect.

Two recent papers are likely to be particularly useful for fitting centiles to normally distributed data (Altman, 1993; Royston & Altman, 1994). Taken together they provide a

very flexible way of modelling the mean and standard deviation of the distribution as it changes with age, using only multiple regression analysis.

I thank Professor Michael Healy for his valuable comments on a previous draft of the paper.

REFERENCES

- Altman, D. G. (1993). Construction of age-related reference centiles using absolute residuals. *Statistics in Medicine* **12**, 917–924.
- Anon. (1993). *Growth Program: software for growth analysis*. Welwyn Garden City: Castlemead Publications.
- Bailey, B. J. R. (1994). Monitoring the heights of prepubertal children by the use of standard charts. *Annals of Human Biology* **21** (in the press).
- Bairagi, R. (1986). On components of variation of estimated weight velocity of children. *Applied Statistics* **35**, 178–182.
- Bairagi, R., Chowdhury, M. K., Kim, Y. J. & Curlin, G. T. (1985). Alternative anthropometric indicators of mortality. *American Journal of Clinical Nutrition* **42**, 296–306.
- Berkey, C. S. (1982). Comparison of two longitudinal growth models for preschool children. *Biometrics* **38**, 221–234.
- Berkey, C. S. & Reed, R. B. (1987). A model for describing normal and abnormal growth in early childhood. *Human Biology* **59**, 973–987.
- Berkey, C. S., Reed, R. B. & Valadian, I. (1983). Longitudinal growth standards for preschool children. *Annals of Human Biology* **10**, 57–67.
- Bland, J. M., Peacock, J. L., Anderson, H. R., Brooke, O. G. & Decurtis, M. (1990). The adjustment of birth weight for very early gestational ages – 2 related problems in statistical analysis. *Applied Statistics* **39**, 229–239.
- Boas, F. (1930). Observations on the growth of children. *Science* **72**, 44–48.
- Bock, R. D. & Thissen, D. (1976). Fitting multi-component models for growth in stature. *Proceedings of the 9th International Biometric Conference* **1**, 431–442.
- Bowditch, H. P. (1891). *The Growth of Children, Studied by Galton's Method of Percentile Grades*. Boston: Wright and Potter.
- Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society B* **26**, 211–252.
- Cameron, N. (1980). Conditional standards for growth in height of British children from 5.0 to 15.99 years of age. *Annals of Human Biology* **7**, 331–337.
- Cameron, N. (1986). The methods of auxological anthropometry. In *Human Growth: a comprehensive treatise*, 2nd edn, Vol. 3, pp. 3–46 [F. Falkner and J. M. Tanner, editors]. New York: Plenum Press.
- Chen, L. C., Chowdhury, A. K. M. A. & Huffman, S. L. (1980). Anthropometric assessment of energy protein malnutrition and subsequent risk of mortality among preschool aged children. *American Journal of Clinical Nutrition* **33**, 1836–1845.
- Chinn, S. (1992). A new method for calculation of height centiles for preadolescent children. *Annals of Human Biology* **19**, 221–232.
- Chinn, S. & Morris, R. W. (1980). Standards of weight-for-height for English children from age 5.0 to 11.0 years. *Annals of Human Biology* **7**, 457–473.
- Chinn, S., Price, C. E. & Rona, R. J. (1989). Need for new reference curves for height. *Archives of Disease in Childhood* **64**, 1545–1553.
- Chinn, S., Rona, R. J., Gulliford, M. C. & Hammond, J. (1992). Weight-for-height in children aged 4–12 years. A new index compared to the normalized body mass index. *European Journal of Clinical Nutrition* **46**, 489–500.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–836.
- Cole, T. J. (1979). A method for assessing age-standardized weight-for-height in children seen cross-sectionally. *Annals of Human Biology* **6**, 249–268.
- Cole, T. J. (1985). A critique of the NCHS weight for height standard. *Human Biology* **57**, 183–196.
- Cole, T. J. (1986). Weight/height^p compared to weight/height² for assessing adiposity in childhood – influence of age and bone age on *p* during puberty. *Annals of Human Biology* **13**, 433–451.
- Cole, T. J. (1988). Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society A* **151**, 385–418.
- Cole, T. J. (1989a). The British, American NCHS and Dutch weight standards compared using the LMS method. *American Journal of Human Biology* **1**, 397–408.
- Cole, T. J. (1989b). Using the LMS method to measure skewness in the NCHS and Dutch national height standards. *Annals of Human Biology* **16**, 407–419.
- Cole, T. J. (1990). The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition* **44**, 45–60.

- Cole, T. J. (1994). Centile charts for both cross-sectional and longitudinal data. *Statistics in Medicine* **13** (in the press).
- Cole, T. J. & Green, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine* **11**, 1305-1319.
- Cole, T. J., Paul, A. A., Eccles, M. & Whitehead, R. G. (1989). The use of a multiple growth standard to highlight the effects of diet and infection on growth. In *Auxology 88: Perspectives in the Science of Growth and Development*, pp. 91-100 [J. M. Tanner, editor]. London: Smith-Gordon.
- Count, E. (1943). Growth pattern of the human physique. *Human Biology* **15**, 1-32.
- Cronk, C., Crocker, A. C., Pueschel, S. M., Shea, A. M., Zackai, E., Pickens, G. & Reed, R. B. (1988). Growth charts for children with Down Syndrome: 1 month to 18 years of age. *Pediatrics* **81**, 102-110.
- Cronk, C. E. & Roche, A. F. (1982). Race- and sex-specific reference data for triceps and subscapular skinfolds and weight/stature. *American Journal of Clinical Nutrition* **35**, 347-354.
- Davies, P. S. W., Day, J. M. E. & Cole, T. J. (1993). Converting Tanner-Whitehouse reference triceps and subscapular skinfold measurements to standard deviation scores. *European Journal of Clinical Nutrition* **47**, 559-566.
- Emery, J. L., Waite, A. J., Carpenter, R. G., Limerick, S. R. & Blake, D. (1985). Apnoea monitors compared with weighing scales for siblings after cot death. *Archives of Disease in Childhood* **60**, 1055-1060.
- Galton, F. (1885). Anthropometric per-centiles. *Nature* **31**, 223-225.
- Goldstein, H. (1986). Sampling for growth studies. In *Human Growth: a comprehensive treatise*, 2nd edn, Vol. 3, pp. 59-78 [F. Falkner and J. M. Tanner, editors]. New York: Plenum Press.
- Goldstein, H. & Tanner, J. M. (1980). Ecological considerations in the creation and the use of child growth standards. *Lancet* **i**, 582-585.
- Graiczer, P. L. & Gentry, E. M. (1981). Measuring children: one reference for all. *Lancet* **ii**, 297-299.
- Grant, J. (1984). *The State of the World's Children 1984*. Oxford: Oxford University Press, for UNICEF.
- Guo, S., Roche, A. F., Baumgartner, R. N., Chumlea, W. C. & Ryan, A. S. (1990a). Kernel regression for smoothing percentile curves: reference data for calf and subscapular skinfold thicknesses in Mexican Americans. *American Journal of Clinical Nutrition* **51**, 908S-916S.
- Guo, S., Roche, A. F., Fomon, S. J., Nelson, S. E., Chumlea, W. C., Rogers, R. R., Baumgartner, R. N., Ziegler, E. E. & Siervogel, R. M. (1991). Reference data on gains in weight and length during the first two years of life. *Journal of Pediatrics* **119**, 355-362.
- Guo, S., Roche, A. F. & Yeung, D. L. (1990b). Monthly growth status from a longitudinal study of Canadian infants. *Canadian Journal of Public Health* **81**, 215-221.
- Guo, S., Siervogel, R. M., Roche, A. F. & Chumlea, W. C. (1992). Mathematical modelling of human growth: a comparative study. *American Journal of Human Biology* **4**, 93-104.
- Hamill, P. V. V., Drizd, T. A., Johnson, C. L., Reed, R. B. & Roche, A. F. (1977). *NCHS Growth Curves for Children, Birth-18 Years*. Washington, DC: National Center for Health Statistics.
- Hammer, L. D., Kraemer, H. C., Wilson, D. M., Ritter, P. L. & Dornbusch, S. M. (1991). Standardized percentile curves of body-mass index for children and adolescents. *American Journal of Diseases of Children* **145**, 259-263.
- Healy, M. J. R. (1962). The effect of age grouping on the distribution of a measurement affected by growth. *American Journal of Physical Anthropology* **20**, 49-50.
- Healy, M. J. R. (1974). Notes on the statistics of growth standards. *Annals of Human Biology* **1**, 41-46.
- Healy, M. J. R. (1988). In the discussion of Cole (1988). *Journal of the Royal Statistical Society A* **151**, 411-412.
- Healy, M. J. R. (1991). The reliability of height and height velocity in the assessment of growth. *Archives of Disease in Childhood* **66**, 1465.
- Healy, M. J. R. (1992). Normalizing transformations for growth standards. *Annals of Human Biology* **19**, 521-526.
- Healy, M. J. R., Rasbash, J. & Yang, M. (1988a). Distribution-free estimation of age-related centiles. *Annals of Human Biology* **15**, 17-22.
- Healy, M. J. R., Yang, M., Tanner, J. M. & Zumrawi, F. Y. (1988b). The use of short-term increments in length to monitor growth in infancy. In *Linear Growth Retardation in Less Developed Countries (Nestlé Nutrition Workshop Series 14)*, pp. 41-55 [J. C. Waterlow, editor]. New York: Raven Press.
- Himes, J. H., Roche, A. F. & Thissen, D. (1981). *Parent-specific Adjustments for Assessment of Recumbent Length and Stature (Monographs in Paediatrics vol. 13)*, pp. 1-88. Basel: Karger.
- Jens, R. M. & Bayley, N. (1937). A mathematical method for studying the growth of a child. *Human Biology* **9**, 556-563.
- Johnston, F. E. & Ouyang, Z. (1991). Choosing appropriate reference data for the anthropometric assessment of nutritional status. In *Anthropometric Assessment of Nutritional Status*, pp. 337-346 [J. H. Himes, editor]. New York: Wiley-Liss.
- Jolicœur, P., Pontier, J. & Abidi, H. (1992). Asymptotic models for the longitudinal growth of human stature. *American Journal of Human Biology* **4**, 461-468.
- Jolicœur, P., Pontier, J., Pernin, M.-O. & Sempé, M. (1988). A lifetime asymptotic growth curve for human height. *Biometrics* **44**, 995-1003.
- Jones, M. C. (1988). In the discussion of Cole (1988). *Journal of the Royal Statistical Society* **151**, 412-413.
- Jones, M. C. & Hall, P. (1990). Mean squared error properties of kernel estimates of regression quantiles. *Statistics and Probability Letters* **10**, 283-289.

- Jordan, J., Ruben, M., Hernandez, J., Bebelagua, A., Tanner, J. M. & Goldstein, H. (1975). The 1972 Cuban national child growth study as an example of population health monitoring: design and methods. *Annals of Human Biology* **2**, 153–171.
- Karlberg, J. (1987). On the modelling of human growth. *Statistics in Medicine* **6**, 185–192.
- Kouchi, M., Mukherjee, D. & Roche, A. F. (1985). Curve fitting for growth in weight during infancy with relationships to adult status, and familial associations of the estimated parameters. *Human Biology* **57**, 245–263.
- Lampl, M., Veldhuis, J. D. & Johnson, M. L. (1992). Saltation and stasis: a model of human growth. *Science* **258**, 801–803.
- Lavoi-Pierre, G. J., Keller, W., Dixon, H., Dustin, J. P. & ten Dam, G. (1983). *Measuring Change in Nutritional Status. Guidelines for Assessing the Nutritional Impact of Supplementary Feeding Programmes for Vulnerable Groups*. Geneva: WHO.
- Lyon, A. J., Preece, M. A. & Grant, D. B. (1985). Growth curve for girls with Turner syndrome. *Archives of Disease in Childhood* **60**, 932–935.
- Marshall, W. A. (1971). Evaluation of growth rate in height over periods of less than one year. *Archives of Disease in Childhood* **46**, 414–420.
- Morley, D. (1973). *Paediatric Priorities in the Developing World*. London: Butterworth.
- Must, A., Dallal, G. E. & Dietz, W. H. (1991). Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht²) and triceps skinfold thickness. *American Journal of Clinical Nutrition* **53**, 839–846.
- Naeraa, R. W. & Nielsen, J. (1990). Standards for growth and final height in Turner's syndrome. *Acta Paediatrica Scandinavica* **79**, 182–190.
- Pan, H. Q., Goldstein, H. & Guo, D. (1992). A two-level cross-sectional model using grafted polynomials. *Annals of Human Biology* **19**, 337–346.
- Pan, H. Q., Goldstein, H. & Yang, Q. (1990). Non-parametric estimation of age-related centiles over wide age ranges. *Annals of Human Biology* **17**, 475–481.
- Pelletier, D. (1991). *Relationships between Child Anthropometry and Mortality in Developing Countries*. Ithaca, NY: Cornell University.
- Piwoz, E. G., Pearson, J. M. & Brown, K. H. (1992). Potential for misclassification of infants' growth increments by using existing reference data. *American Journal of Clinical Nutrition* **56**, 58–64.
- Preece, M. A. & Baines, M. J. (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology* **5**, 1–24.
- Ranke, M. B. & Guilbaud, O. (1990). Growth response in prepubertal children with idiopathic growth hormone deficiency during the first year of treatment with human growth hormone. Analysis of the Kabi international growth study. *Acta Paediatrica Scandinavica* Suppl. 370, 122–130.
- Ranke, M. B., Pflüger, H., Rosendahl, W., Stubbe, P., Enders, H., Bierich, J. R. & Majewski, F. (1983). Turner's syndrome: spontaneous growth in 150 cases and review of the literature. *European Journal of Pediatrics* **141**, 81–88.
- Roche, A. F., Guo, S. & Moore, W. M. (1989). Weight and recumbent length from 1 to 12 mo of age: reference data for 1-mo increments. *American Journal of Clinical Nutrition* **49**, 599–607.
- Roede, M. J. & Van Wieringen, J. C. (1985). Growth diagrams 1980. Netherlands third nation-wide survey. *Tijdschrift voor Sociale Gezondheidszorg* **63** (Suppl.), 1–34.
- Rolland-Cachera, M. F., Cole, T. J., Sempé, M., Tichet, J., Rossignol, C. & Charraud, A. (1991). Body mass index variations: centiles from birth to 87 years. *European Journal of Clinical Nutrition* **45**, 13–21.
- Rolland-Cachera, M. F., Deheeger, M., Bellisle, F., Sempé, M., Guillaud-Bataille, M. & Patois, E. (1984). Adiposity rebound in children: a simple indicator for predicting obesity. *American Journal of Clinical Nutrition* **39**, 129–135.
- Rolland-Cachera, M. F., Sempé, M., Guillaud-Bataille, M., Patois, E., Péquignot-Guggenbuhl, F. & Fautrad, V. (1982). Adiposity indices in children. *American Journal of Clinical Nutrition* **36**, 178–184.
- Rossiter, J. E. (1991). Calculating centile curves using kernel density estimation methods with application to infant kidney lengths. *Statistics in Medicine* **10**, 1693–1701.
- Rowland, M. G. M., Cole, T. J. & Whitehead, R. G. (1977). A quantitative study into the role of infection in determining nutritional status in Gambian village children. *British Journal of Nutrition* **37**, 441–450.
- Royston, P. & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics* **43**, (in the press).
- Sempé, M., Pédrón, G. & Roy-Pernot, M. (1979). *Auxologie: méthode et séquences*. Paris: Theraplix.
- Shohji, T. & Sasaki, H. (1987). Individual growth of stature of Japanese. *Growth* **51**, 432–450.
- Simondon, K. B., Simondon, F., Delpeuch, F. & Cornu, A. (1992). Comparative study of five growth models applied to weight data from Congolese infants between birth and 13 months of age. *American Journal of Human Biology* **4**, 327–335.
- Smith, D. W., Truog, W., Rogers, J. E., Greitzer, L. J., Skinner, A. L., McCann, J. J. & Harvey, M. A. S. (1976). Shifting linear growth during infancy: illustration of genetic factors in growth from fetal life through infancy. *Journal of Pediatrics* **89**, 225–230.
- Stützle, W., Gasser, T., Molinari, L., Largo, R. H., Prader, A. & Huber, P. J. (1980). Shape-invariant modelling of human growth. *Annals of Human Biology* **7**, 507–528.
- Tanner, J. M. (1962). *Growth at Adolescence, with a general consideration of the effects of hereditary and environmental factors upon growth and maturation from birth to maturity*, 2nd edn. Oxford: Blackwell.

- Tanner, J. M. (1981). *A History of the Study of Human Growth*. Cambridge: Cambridge University Press.
- Tanner, J. M. & Davies, P. S. W. (1985). Clinical longitudinal standards for height and height velocity for North American children. *Journal of Pediatrics* **107**, 317–329.
- Tanner, J. M., Goldstein, H. & Whitehouse, R. H. (1970). Standards for children's height at ages 2–9 years allowing for height of parents. *Archives of Disease in Childhood* **45**, 755–762.
- Tanner, J. M., Lejarraga, H. & Healy, M. J. R. (1972). Within-family standards for birth-weight. *Lancet* **ii**, 1314–1315.
- Tanner, J. M., Whitehouse, R. H. & Takaishi, M. (1966). Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965 Parts I and II. *Archives of Disease in Childhood* **41**, 454–471, 613–635.
- Thompson, M. L. & Theron, G. B. (1990). Maximum likelihood estimation of reference centiles. *Statistics in Medicine* **9**, 539–548.
- Toole, M. J. & Malkki, R. M. (1992). Famine-affected, refugee, and displaced populations: recommendations for public health issues. *Morbidity and Mortality Weekly Report* **41** (Supplement RR-13), 1–25.
- Van Lerberghe, W. (1988). Growth, infection and mortality: is growth monitoring an efficient screening instrument? In *Auxology 88: Perspectives in the Science of Growth and Development*, pp. 101–110 [J. M. Tanner, editor]. London: Smith-Gordon.
- Vaughan, V. C. & McKay, R. J. (1975). *Nelson Textbook of Pediatrics*, 10th edn. Philadelphia, PA: Saunders.
- Voss, L. D., Bailey, B. J. R., Cumming, K., Wilkin, T. J. & Betts, P. R. (1990). The reliability of height measurement (the Wessex growth study). *Archives of Disease in Childhood* **65**, 1340–1344.
- Voss, L. D., Mulligan, J., Betts, P. R. & Wilkin, T. J. (1992). Poor growth in school entrants as an index of organic disease: the Wessex growth study. *British Medical Journal* **305**, 1400–1402.
- Voss, L. D., Wilkin, T. J. & Betts, P. R. (1987). Do we need new growth charts? *Lancet* **ii**, 447–448.
- Waterlow, J. C. (1972). Classification and definition of protein-calorie malnutrition. *British Medical Journal* **iii**, 566–569.
- Waterlow, J. C., Buzina, R., Keller, W., Lane, J. M., Nichaman, M. Z. & Tanner, J. M. (1977). Presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. *Bulletin of the World Health Organization* **55**, 489–498.
- Whitehead, R. G. & Paul, A. A. (1984). Growth charts and the assessment of infant feeding practices in the western world and in developing countries. *Early Human Development* **9**, 187–207.
- WHO (1986). Use and interpretation of anthropometric indicators of nutritional status. *Bulletin of the World Health Organization* **64**, 929–941.
- Wright, C. M., Matthews, J. N. S., Waterston, A. & Aynsley-Green, A. (1994). What is a normal rate of weight gain in infancy? *Acta Paediatrica* (in the press).