Assessment: National and International Reference Standards

Timothy J. Cole

Center for Pediatric Epidemiology and Biostatistics, Institute of Child Health, London, England, UK

Child obesity is a serious and increasing problem worldwide. It has important psychosocial and physical effects during childhood and later, in adult life. There is an important need to quantify the scale of the problem, for three reasons: to establish its prevalence in different parts of the world, to monitor the trends in prevalence over time, and to test the effectiveness of interventions introduced to address the problem.

DEFINITION OF CHILDHOOD OBESITY

Obesity is very simple to define qualitatively. It is an excess of body fat. But a quantitative definition requires wide agreement about how to measure body fat, and what cutoff points to use to separate the obese from the nonobese. An extra requirement in childhood is to link the cutoffs to the child’s age and sex, because child growth ensures that body size, and hence the amount of body fat, changes with age in a way that is quite unrelated to obesity. So the definition of obesity throws up two requirements: (a) a way to measure body fat and (b) a set of age- and sex-specific cutoff points. A third requirement is that these two requirements are widely recognized, so that there is a degree of standardization in the definition.

MEASUREMENT OF BODY FAT

Another name for body fat is adipose tissue, and adiposity is the state of having body fat. Compared with obesity, this is a neutral term to describe the spectrum of fatness from obese through very thin. Adipose tissue is spread round the body, some of it subcutaneous and some internal. Subcutaneous fat can be accessed through the skin and so can be measured to reasonable accuracy using skinfold calipers. But internal fat cannot be reached in this way and it requires more expensive whole-body research instruments to measure it. The amount of central body fat—that is, at waist level, both internal and subcutaneous—can be estimated by measuring the waist circumference, though this also involves other internal organs, such as the liver.
There are more complex methods—for example, dual-energy x-ray absorptiometry (DEXA) or isotope dilution—that provide whole-body measures of adipose tissue, either directly or indirectly, but for population use they are impractical, in terms of both cost and availability. This constrains useful population measures of body fat to anthropometry—skinfold thickness measurements, circumferences, weight, and height.

**Body Mass and Body Fat**

Unlike skinfolds and circumferences, weight and height are whole-body measures. Weight, though related to fat, does not measure it directly, whereas height has little relation to fat. Weight and height are fairly highly correlated, with a correlation coefficient of around 0.7 during childhood; weight adjusted for height has long been considered to be a simple and convenient index of body shape, independent of body size (1).

This view first developed in the field of malnutrition, where weight adjusted for height (or weight-for-height) was used as a measure of wasting—that is, it indicated the amount of body tissue that had been lost because of malnutrition. There was no suggestion that the lost tissue was purely fat; indeed, it was recognized that much of it was fat-free tissue. For this reason malnutrition should not be viewed simply as the opposite of obesity; it is not simply a deficit of body fat. For the same reason, weight-for-height used to assess malnutrition should not be viewed simply as the mirror image of weight-for-height used to assess obesity. The relation between weight-for-height and body fat is important for obesity, but less so for malnutrition.

**Weight–Stature Indices**

The term weight-for-height has been in use for 40 years or more, yet it can be a confusing concept. Many ways exist for statistically adjusting weight-for-height, some better than others, and they all involve slightly different assumptions. A brief description of some of the issues is given here, but for a more complete discussion see Cole (1). It is easier to consider adults and children separately, as age is an important extra confounding variable in children that is less relevant for adults.

**Adults**

The simplest way to adjust weight-for-height is to identify a reference population, split it into a series of height classes by sex, and average the weights in each height class. An individual’s weight can then be compared with this reference weight for his or her height and sex. But although this approach is simple to do, it is inefficient. The mean weights in the different height cells tend to increase smoothly with height, so the means can be smoothed across cells. This allows the mean weights to be represented as a smooth curve drawn on a chart as a function of height. This is traditionally known as the weight-for-height chart. Median rather than mean weight may alternatively be used.
If the weight–height trend is linear, then the linear regression of weight on height provides a compact formula relating expected weight to height, where expected weight is of the form \([a + b \times \text{height}]\) and \(a\) and \(b\) are estimated regression coefficients. This leads to a difference index of the form

\[
\text{weight} - \text{expected weight-for-height}
\]

which is a direct measure of over- or underweight, given height. However, this form of index is less intuitively appealing than the ratio index:

\[
\frac{\text{weight}}{\text{expected weight-for-height}}
\]

which is more commonly used.

Linear regression presents two problems in this context: heteroscedasticity and skewness. The first of these is the phenomenon of nonconstant variability of weight; its standard deviation tends to increase as mean weight increases. The second problem, skewness, is the upper tail of the weight distribution tending to be longer and more extended than the lower tail. This is most apparent on a chart where the upper centiles are further apart than the lower centiles. The presence of skewness or heteroscedasticity reduces the validity of the regression equation, and the correct procedure is to find a transformation of weight that will remove them. A popular and effective transformation is to take logarithms, so the regression equation involves log weight and (by symmetry) log height rather than weight and height. This leads to the regression equation \(\log \text{weight} = a + p \times \log \text{height}\), and a difference index of the form

\[
\log \text{weight} - p \times \log \text{height}
\]

When antilogged this gives the ratio index: \(\text{weight/height}^p\).

Two important corollaries arise from this analysis. The first is that the log transform, by converting a difference index to a ratio index, provides a statistical justification for the ratio index. The second relates to the power of height \(p\). Using regression analysis means that the value of \(p\) tends to be less than the number 3 implied using dimensional arguments (i.e., the argument that weight is proportional to volume, which is proportional to length\(^3\)).

In adults \(p\) is close to 2 in men, and somewhat smaller in women. Despite this sexual dimorphism, the value 2 has been preferred, and the index weight/height\(^2\) is now in universal use for assessing overweight and obesity in adults, both men and women (2). It is often called Quetelet’s index, after the Belgian mathematician Adolphe Quetelet, who first observed that weight goes as the square of height, but more recently it has become known as the body mass index (BMI) (3).

**Children**

Moving from adults to children, age becomes a third dimension, in addition to weight and height, that needs to be accounted for. Weight adjusted for height may or may not be adjusted for age as well. The weight-for-height chart described earlier, which
ignores the child's age, is useful to monitor malnutrition in areas of the world where age is not recorded. Cutoff points for malnutrition have in the past used the ratio index:

\[
\text{weight/expected weight-for-height} \times 100,
\]

where less than 80% of expected weight is the definition of wasting (4).

However, in many parts of the world the child's age is known, and ignoring it is inefficient and leads to bias. Weight averaged in narrow height classes can be extended to include tabulation by age as well—that is, the height classes are defined within age classes, though this requires a very substantial sample size. Baldwin (5) used this approach as long ago as 1925 to provide height-, age-, and sex-specific values for weight in American children.

Within each age class the relation between weight and height can be summarized using log-log regression, as described earlier, which leads to the index weight/height\(^2\), where \(p\) depends on age. The question then is, "Does \(p\) change much with age or is it fairly constant?" Rolland-Cachera et al. (6) showed almost 20 years ago that height\(^2\) was appropriate in early childhood but that during puberty the relation was steeper and height\(^3\) was better. Three years earlier, Cole (7) had come to the same conclusion using a different statistical approach, expressing weight and height as weight-for-age and height-for-age, respectively—that is, dividing each measurement by a reference value for the child's age and sex. The regression of log weight-for-age on log height-for-age adjusts for age and estimates the optimal height power \(p\) simultaneously.

This age dependence in the value of the height power has led to some soul searching. On the one hand, people have argued on pragmatic grounds that weight/height\(^2\) should be used throughout childhood. This would mean that during puberty the index does not adjust wholly for the effects of height, and taller children will tend to have a larger index—that is, taller children will appear fatter. The same does not hold earlier in childhood, where the index is broadly uncorrelated with height.

On the other hand, there is a feeling that the index should be uncorrelated with height at all ages, even though this involves at least two distinct indices, weight/height\(^2\) and weight/height\(^3\). A more purist approach still is to allow the height power to change smoothly with age (8), which complicates the issue further.

It is useful at this point to ask what the index is for. Is it to measure weight-for-height or is it to identify obesity? If the former, then we are obliged to use the age-related height power, as otherwise the index will not adjust properly for height at all ages. But if our aim is to measure obesity (which it is), then we require the index to measure body fat as best it can.

Before puberty both approaches identify the BMI as the optimal index; it is uncorrelated with height and maximally correlated with body fat. During puberty an association develops between body fat and height; fatter children tend also to be taller, so the optimal index of weight-for-height, weight/height\(^3\) needs to be adjusted further for height. This again leads to the BMI. So for assessing obesity BMI is the optimal index throughout childhood and adolescence, as it is during adulthood.
Weight and Height Centiles and Z Scores

Anthropometry is often expressed in the form of a standard deviation score, SD score, or z score. This is the number of standard deviations that the individual child’s measurement is above or below the median of the distribution for age and sex. So a height z score of 0 indicates median height for age, whereas +2 z scores is two standard deviations above the mean and corresponds to the 98th centile. Based on the reference population, the z score has a mean of 0 and a standard deviation of 1 and is normally distributed.

Until recently, the z score calculation has required the measurement to be normally distributed, so that weight, BMI, or skinfold thickness—which each have markedly skewed distributions—could not be expressed in z score form. However, this has now become possible using a power transformation of the measurement, so that z scores for weight, height, and BMI are now available. The British and American references are both constructed on this basis, a method known as the LMS method (see later).

The availability of weight z score and height z score provides a new way of defining weight-for-height, using the regression of weight z score on height z score. It is a logical extension of the two regression approaches described earlier: log weight on log height, and log weight-for-age on log height-for-age. But the z score does not need a log transformation, as it is already normally distributed. This regression leads to a difference index of the following form:

\[ \text{weight z score} = b \times \text{height z score} \]

where the coefficient \( b \) lies in the range 0.5 to 0.7, depending on age and sex (unpublished data).

This apparently complex formula links directly to the informal process that pediatricians employ to assess a child’s fatness from weight and height. They look at the child’s weight and height charts, read off the corresponding centiles, and compare the child’s deviations from the median. If the two deviations are similar they conclude that the child’s weight is appropriate for his or her height. This assessment is a form of difference index that treats the weight and height centiles as equally important.

Statistically speaking, it is more appropriate to compare z scores than centiles, for two reasons: centiles are on a nonlinear scale that is elongated in the tails of the distribution, and in any case z scores provide a direct measure of the deviation from the median. So the informal procedure is equivalent to treating the difference in z score between weight and height as a measure of weight-for-height. But the preceding formula shows that the two z scores should not be given equal importance—height is only half to two-thirds as important as weight.

So the informal process that compares weight and height centiles needs modifying. For children of average fatness, the weight z score is roughly 50% greater than the height z score. If it is appreciably more than 50%, this is a sign of fatness; if it is appreciably less, this indicates thinness. In the simplest case, the child’s weight and height are both on the median for age, and this represents average fatness by definition.

The same approach can be extended to BMI expressed as a z score, which it turns out can also be expressed to high accuracy as a difference index of the form \[ \text{weight z score} = b \times \text{height z score} \] (31).
Body Mass Index: Pros and Cons

The body mass index has now been embraced wholeheartedly for use in children. Since the early 1990s BMI growth charts have been published for several countries, providing population centile curves for BMI by age and sex. These charts allow the individual child’s BMI to be expressed as a centile, and where the centile is too high or there is appreciable upward centile crossing over time, the child is at increased risk of obesity.

There are several advantages to the BMI for assessing obesity: It is based on measurements that are simple and cheap to make; it provides an assessment based on the whole body; it is highly correlated with body fat and broadly uncorrelated with height; and it is the index used in adults. Against these advantages there is one major disadvantage—it cannot distinguish between fat mass and fat-free mass. Consider two children matched for age, weight, and height, but one fat and the other lean. Because their weights and heights are the same, they have the same BMI yet they differ considerably in body composition. The excess fat in the one child weighs the same as the extra muscle in the other child, and the BMI is unable to distinguish between them.

This is potentially a serious problem for monitoring trends in obesity. If the trend to increasing fatness over time is matched by a corresponding decrease in muscle mass, the trend will be invisible to the BMI. This switch from fat-free to fat mass is tending to happen, as one of the main risk factors for obesity is sedentary behavior leading to a lack of fitness and reduced muscle mass. The problem does not arise with obesity indices based directly on body fat (e.g., skinfold thickness). Thus Flegal (9) has shown that trends in obesity over time in American adolescents are much more obvious when based on changes in triceps skinfold than when based on changes in BMI. This means that apparent trends in BMI are likely to underestimate the scale of the underlying problem.

REFERENCES AND STANDARDS

Growth charts based on reference standards have been mentioned several times in this chapter, but here they are discussed in more detail. The first issue is what to call them—references or standards. References are normative; that is, they are based on a reference population representative of some geographic or cultural group, whereas standards are prescriptive—they are based on a population selected on the basis of health. The implication is that standards represent optimal growth in some sense, whereas references indicate typical growth.

A practical problem with standards lies in the inclusion and exclusion criteria needed to define the reference population. The criteria are inevitably arbitrary (e.g., should severely asthmatic children be excluded, and if so how is severe asthma defined?), and this weakens the objectivity of the process. In practice, most growth charts are references rather than standards, constructed to be representative of one country. The term growth standards is, strictly speaking, incorrect in this context.
Purpose

Growth charts based on growth references are designed to assess measurements at a single moment in time, but in practice they are used to monitor growth over extended periods. This distinction is important to understand the strengths and weaknesses of the growth chart. The reference data underlying the chart usually consist of single measurements from a large sample of children covering the required age range, obtained from a cross-sectional anthropometric survey. Such a dataset contains no information about growth over time, as each child contributes just one measurement. So the term growth chart is, strictly speaking, inappropriate—the chart measures body size at different ages, not growth.

But doesn’t everyone use growth charts to assess growth? Yes, but only in a limited sense. The assumption is that once they are past infancy, children grow along their own chosen centile on the chart. Normal growth is represented by a constant centile, and abnormal growth is seen as centile crossing, either up or down. This is fine so far as it goes, but in practice no child stays on exactly the same centile. Inevitably, some centile crossing occurs from age to age. The key question is, “How much centile crossing is reasonable, and at what point does it become excessive?” This is a question that the conventional growth or “distance” chart fails to answer. What is needed here is a “velocity” chart, which quantifies the change in the measurement (or its z score) over time (10). It is also possible to represent velocity information on the distance chart by adding extra curves with slopes that indicate a given rate of centile crossing (11).

The degree of centile crossing that occurs depends on the correlation among measurements at different ages. Height, for example, tracks very strongly, with children tending to stay close to their chosen centile in late infancy until they reach puberty. The year-on-year correlation for height is about 0.98 in mid-childhood (12). For weight the correlation is slightly lower, and the amount of tracking is less. For BMI, which is weight adjusted for height, the tracking is even less, as the stabilizing effect of height has been adjusted out. So a child’s BMI centile is likely to change over time considerably more than for height.

The correlation of 0.98 seems very high, as indeed it is, but the amount of centile crossing over time depends not on the correlation $r$ itself, but on the function $\sqrt{1 - r^2}$ of $r$, which is the standard deviation of the change in z score over the period (11). So if $r = 0.98$ then $\sqrt{1 - r^2} = 0.20$, and 95% of subjects followed over a year will cross centiles by less than $2 \times 0.20 = 0.40$ z scores up or down, about two-thirds of the width between two centile curves. (This channel width is typically 0.67 z scores 13.) For BMI the correlation $r$ is about 0.94 (Rudolf M, personal communication), so $\sqrt{1 - r^2} = 0.34$ and the 95% confidence interval is $2 \times 0.34 = 0.68$ z scores or one channel width, 50% more than for height. A modest reduction in the correlation leads to a dramatic increase in the amount of centile crossing over 1 year.

Over a 2-year period the correlation for BMI is roughly $0.94 \times 0.94 = 0.89$, which is one-third more centile crossing than for 1 year. And for each successive year that
BMI is monitored, the correlation shrinks by the factor 0.94 and the degree of centile crossing increases accordingly.

This is a weakness of the BMI chart, as centile crossing is a more sensitive risk factor for later obesity than a high centile at any one moment in time. There is as yet no information about the limits of centile crossing to guide the use of the BMI chart. This is an area requiring further research.

**Construction**

So far we have discussed growth centile charts without considering where they come from. After the reference sample has been identified and the measurements collected, the statistical construction of the centile curves is the logical next stage. The methodology of centile curve construction has developed considerably in the last 15 years and has focused on two distinct approaches. The first, usually termed quantile regression (14), estimates the shapes of a series of prespecified centile curves (e.g., the third, 50th, and 97th centiles), using only the rankings of the measurements at each age in the region of the particular centile. Take the third centile, for example: For each of a series of narrow age groups the data are sorted into order and the cutoff point is identified, which separates the smallest 3% from the remainder. These cutoff points are smoothed across ages to provide the third centile curve. The same process leads to the median and 97th centile curves.

An important property of quantile regression is that the shapes of individual curves do not affect the shapes of neighboring curves. There is no guarantee that they will be similar in shape, and in principle the curves can touch or even cross. Quantile regression is most useful when one single centile is of interest (e.g., the 99th centile for wind speed over a period of time). But when the requirement is for a set of several centile curves rather than just one, as here, quantile regression is inefficient. The curves need to be similar in shape and spaced a suitable distance apart. It ought to be possible for curves to "borrow strength" from their neighbors during the estimation process.

This summarizes the other major approach to centile curve construction, which assumes some form of underlying frequency distribution at each age. The distribution ensures that the centiles are appropriately spaced, and equally that the centile curves plotted against age are of similar shape. The frequency distribution may—indeed will—change with age, but as long as the change is smooth and gradual, the centile curves will themselves be smooth.

The simplest distribution is the normal (Gaussian) distribution, and certain measurements (e.g., height) are close to normally distributed. This means that the distribution of height by age can be summarized in terms of the age-specific mean and SD, and these in turn define the required centiles of the distribution. Assume that \( z_\alpha \) is the normal equivalent deviate corresponding to a tail area of \( \alpha \), so that 100\( \alpha \)% of the distribution is to the left of \( z_\alpha \). The corresponding centile is given by the following:

\[
\text{Centile } 100\alpha = \text{Mean} + \text{SD} \times z_\alpha
\]

Any required centile can be obtained by setting \( z_\alpha \) appropriately. For example, \( z_\alpha = -1.88 \) corresponds to the 3rd centile and 0.67 to the 75th centile.
Commonly, the standard deviation of the distribution increases in step with the mean, as can be seen on many growth charts—the centiles expand as the mean increases. Another way of describing the standard deviation is to express it as a fraction of the mean. This is known as the coefficient of variation, or CV, where $\text{CV} = \text{SD/mean}$. The advantage of the CV is that with child anthropometry it is less dependent on age than the SD, and so is simpler to estimate. The corresponding formula for centile $100\alpha$ is

$$\text{Centile } 100\alpha = \text{Mean} (1 + \text{CV} \times z_\alpha) \tag{[1]}$$

The elegance of this approach lies in its simplicity and parsimony. It depends only on two variables, the mean and CV, each of which can be represented as smooth curves plotted against age; it produces any required centiles, depending only on the choice of $z_\alpha$, and it is much more efficient than quantile regression (15). Its main problem in the past has been that many anthropometric variables are not normally distributed, which has ruled it out.

Statisticians often have to deal with data that are not normally distributed, where one tail of the distribution (usually the right) is longer than the other—known as skewness. They commonly use a transformation (e.g., taking logarithms) to make them more normally distributed. The transformation alters the shape of the distribution by stretching the lower half and shrinking the upper half. In this way it removes the skewness and “pulls” the distribution closer to normality. The log transformation is only one of a whole family of power transformations (16) that allow the stretching-shrinking process to be tailor made to the individual distribution, to ensure that after transformation the two tails are of equal length. The transformation involves creating a new variable $Y$ from the original variable $X$ such that $Y = X^L$. The log transformation is a special case where $L = 0$, while other popular transformations are the square root ($L = 0.5$) and the reciprocal ($L = -1$).

To ensure that the transformation removes the skewness in the distribution it needs to take nonintegral values, and it needs to change smoothly with age. This is because the degree of skewness itself changes with age. So the power $L$ can be thought of as a third age-varying variable, after the mean and the CV, which is represented as a smooth curve plotted against age.

The three curves allow centiles to be constructed for any anthropometric data with a normal or a skew normal distribution. (Skew normal means that a power transformation makes the distribution normal.) In practice this covers virtually all anthropometry. It has been named the LMS method, the letters of the name indicating the three underlying curves: $L$ the power, $M$ the mean (actually the median), and $S$ the CV (17,18). $M$ is the median rather than the mean because it is calculated as the mean on the normal transformed scale and back-transformed. The mean of a normal distribution is also the median, and on back-transformation it retains its position in the middle of the distribution. The formula for centile $100\alpha$ is

$$\text{Centile } 100\alpha = M(1 + L \times S \times z_\alpha)^{1/L} \tag{[2]}$$

where the values for $L$, $M$, and $S$ are read off the corresponding curves for a child of given age and sex. The centile curve is obtained by calculating centiles over a range
FIG. 1. Body mass index LMS curves by sex for British children, 1990 (23).
of ages. Substituting $L = 1$ into this formula gives the simpler case $M(1 + S \times z_a)$ for normally distributed data, as shown in formula (1).

Other approaches to centile construction have used the same principle of a third curve to represent skewness, though with different parameterizations (19–21). Apart from the parameterization another issue is the form of the smooth curves that are to be fitted—polynomial, cubic spline, or kernel smoother. The LMS method in its most recent form (22) uses cubic splines, which allow for quite complicated curve shapes but at the cost of requiring a table of values by age rather than providing a simple formula.

An important advantage of all these distribution-based approaches is that anthropometric measurements can be converted with full accuracy to $z$ scores. With the LMS method, for example, a measurement $\text{Anth}$ is converted to a $z$ score $Z$ with the following formula:

$$Z = \frac{(\text{Anth}/M)^L - 1}{L \times S}$$

where $L$, $M$, and $S$ are read from the curves and are appropriate for the child’s age and sex.

As an example of the LMS method, Fig. 1 shows the fitted $L$, $M$, and $S$ curves for the British BMI reference by sex (23). The $M$ curves show median BMI for boys and girls, with the familiar rise then fall during the first year, the adiposity rebound at 6 years, and the subsequent rise to adulthood. The $S$ curve is the coefficient of variation of BMI, which is near 0.08, or 8%, before puberty and then jumps to 12% during and after puberty. The $L$ curve shows the changing degree of skewness. Already at birth the distribution is sufficiently skewed to require a log transformation ($L = 0$), but soon afterward the skewness increases and a reciprocal transformation ($L = -1$) becomes necessary. The reciprocal of BMI is height$^2$/weight, and so surprisingly this upside-down index is close to normally distributed. There are two obvious sex differences in the curves: During puberty, median BMI increases faster and flattens earlier in girls than in boys; and the variability of BMI is consistently greater in girls than in boys from as early as 2 years of age.

The centile curves that result from the LMS curves of Fig. 1 are shown in Fig. 2. The current British growth charts use centiles that are equally spaced, two-thirds of a unit apart on the $z$ score scale. This leads to the following (approximate) centiles as seen in Fig. 2: 2, 9, 25, 50, 75, 91, and 98. Each centile curve is the same shape as the $M$ curve, but the centiles splay out with increasing age, and the spacings between them are much greater above the median than below, indicating the presence of appreciable skewness. The age of the second rise in adiposity, known as the adiposity rebound (24), occurs 3 to 4 years earlier on the 98th than on the 2nd centile, suggesting that an earlier adiposity rebound is associated with greater fatness.

**CHOICE OF CUTOFF POINTS**

**Risk-Based Cutoff Point**

The definition of obesity requires a cutoff for BMI, with obesity deemed to be present above the cutoff and absent below it. A second, less extreme cutoff defining
FIG. 2. Body mass index centile curves by sex for British children, 1990 (23).
overweight or the risk of overweight may also be required. But how is this cutoff to be identified?

The most obvious approach is to place it at the point on the BMI distribution that identifies the small number of individuals who, because of their obesity, are at risk of future ill health. But the problem is that this point cannot be identified with any confidence—the link between current fatness in childhood or adolescence and a later adverse outcome (e.g., adult hypertension, diabetes, or cardiovascular disease) is simply too weak to be useful. The risk increases steadily with increasing fatness, but there is no obvious “shoulder” on the plot of risk versus BMI that would indicate a demarcation between low risk and high risk. And even if there were a clear link between child/adolescent fatness and later morbidity/mortality, it would be mediated through adult fatness, which is known to be linked to both. Is a fat child who stays fat at greater risk than a thin child who becomes fat? Or, to put it another way, is the increase in fatness from child to adult a risk factor in its own right?

There are two further practical difficulties with identifying a cutoff based on risk. Cutoff points are required for children of all ages, so a separate cutoff needs to be developed for each age. This requires far more detailed information about the link between current fatness and later outcome than is now available. And the final problem is that the etiology of obesity may have changed in recent years—its steeply increasing prevalence may indicate a different form of the condition from that in the past. If true, this means that the available retrospective evidence linking fatness and outcome may be irrelevant for future predictions.

The conclusion is that for all these reasons, child cutoff points for BMI based on risk are currently not useful. Whether this will remain so in the future is not clear.

National BMI Cutoff Point

The obvious alternative to a risk-based cutoff point is a cutoff relying on BMI centiles. This is population based to ensure that a fixed percentage of the reference population lies above the cutoff point at all ages (e.g., 5% above the 95th centile). Though this is simple and convenient (so long as a suitable BMI centile chart exists), it makes the unlikely assumption that the degree of overweight/obesity is the same at all ages. In practice, this assumption is untestable, as no gold standard exists to measure the prevalence of obesity at each age. In the continuing absence of such a gold standard the simpler definition must suffice.

But even this decision does not solve all the problems. We still have to decide what centile to use for the cutoff point. Bearing in mind that the choice of centile defines the prevalence of obesity (i.e., 5% for the 95th centile), the decision is immediately seen to be as much political as clinical. Barlow and Dietz (25) have provided a clear lead in the United States for the use of the 85th and 95th centiles for overweight and obesity, respectively, but their recommendation has been far from universally accepted. A wide range of centiles has been proposed for obesity, overweight, or risk of overweight (e.g., the 85th, 91st, 95th, 97th, 97.5th, and 98th centiles). These provide a sevenfold difference in overweight/obesity prevalence, from 2% to 15%, an
enormous range. So long as child obesity studies use their own centile cutoff points and their own BMI reference charts, the prevalences they report will not be comparable. Any statements about the prevalence of overweight or obesity in particular groups are meaningless unless they are accompanied by a description of the definition used.

International BMI Centile Chart

The lack of standardization in the definition of childhood overweight and obesity has been a serious obstacle to obesity research in recent years. As a simple example, the section on child obesity in the WHO obesity report (26) started with the admission that there was no agreed definition of child obesity, which undermined everything else that followed. The problem with establishing a standardized definition is that it must be seen to be international in scope. In the past this has not been a problem: the WHO international growth reference, for example, based on American data, has been widely used over the last 20 years (27). However, the political climate has now changed, and a single-country reference like this would no longer be acceptable in the same way.

An international BMI centile chart could be achieved by pooling reference data from several countries, which would avoid the single-country problem. But this raises other problems of its own: how would the countries be chosen, and would their inclusion imply that they were healthier in some way than other countries? The choice of countries would define the level of obesity that was acceptable—higher for countries from the developed world than from the developing world.

But even if the source of the reference data could be agreed, there would still be the problem of what cutoff point to use. Here again, past experience may no longer be relevant, and a different approach is needed. A workshop held in 1997 by the International Obesity Task Force addressed just this question and came up with a novel proposal that has been widely welcomed (28). It links the child BMI centile cutoff points to the cutoff points used in adults—that is, 25 kg/m² for overweight and 30 kg/m² for obesity—which are universally recognized. It also creates this link using the LMS method, as described earlier, to construct the centile chart.

Figure 3 illustrates the process with the British 1990 reference. The LMS method allows any centile curves to be drawn, once the individual L, M, and S curves have been derived (Fig. 1). A particular centile is set by the choice of the normal equivalent deviate \( z_c \) in the formula [(1) and (2)]. To obtain the value of \( z_c \) for a centile curve passing through, say, 25 kg/m² at 18 years, we substitute \( Anth = 25 \) into formula [3], and this gives the \( z \) score (\( Z \)), which is the required value of \( z_c \) in formula [2]. Substituting it into formula [2], we get values of the corresponding centile curve over the whole age range, which by definition passes through BMI 25 at 18 years. Figure 3 shows the resulting curves for cutoff points 25 and 30 at 18 years in British children, corresponding roughly to the 90th and 99th centiles, respectively.

This same process can be carried out on other datasets for which LMS curves exist, and they each lead to a pair of centile curves like those for British boys and girls.
FIG. 3. Body mass index percentile curves for British children 1990 (23) with extra centile curves added that pass through the adult cutoff points of 20 and 30 kg/m² at 18 years.
FIG. 4. Body mass index centile curves by sex for British children 1990 (23), with the IOTF overweight and obesity cutoff points (28) superimposed, that pass through the adult cutoff points of 25 and 30 kg/m² at 18 years.
in Fig. 3. Cole et al. (28) applied the process to six large nationally representative child BMI surveys from Brazil, Britain, Hong Kong, the Netherlands, Singapore, and the United States. The six country overweight centile curves (all passing through BMI 25 at 18 years) and the six country obesity curves (BMI 30 similarly) were then averaged at each age to give two composite centile curves by sex, and they are shown in Fig. 4 superimposed on the British 1990 reference. Comparison with Fig. 3 shows that the cutoff points are similar to those for British children (not surprisingly, as Britain is one of the six countries), though the International Obesity Task Force (IOTF) cutoff points are slightly higher during puberty.

The cutoff points are defined between the ages of 2 and 18 years and are known as the IOTF overweight and obesity cutoffs. (See the British Medical Journal website version of reference 28 for a table of the cutoff points by age and sex.) At the time of writing these cutoff points are relatively novel, so that few studies have made use of them so far.

One example is from Chinn and Rona (29), who have used the IOTF cutoff points to track changes in overweight and obesity from 1974 to 1994 in English and Scottish children of primary school age (4 to 11 years). From 1974 to 1984 the prevalence rates in England fell slightly, while those in Scotland rose slightly. From 1984 to 1994 there were steep rises in both countries, reaching 6% to 9% in the 9- to 11-year-olds. The prevalence of overweight in 1994 was 9% and 14% in boys and girls, respectively, and for obesity it was 2% and 3%. In the future, other studies should show the degree to which these prevalences vary by time and place over the world.

No Cutoff Point at All

So far we have been talking about the need for cutoff points. However, it is worth emphasizing that cutoff points are not always necessary and that the whole BMI distribution can provide useful information about obesity in the population. Mean BMI, for example, tends to be raised in populations where the rate of obesity is high, whatever cutoff is used to define obesity. Similarly, the standard deviation of BMI is increased in fat populations, because BMI in the fattest children tends to outstrip the mean and increases the overall variability. Cutoff points are an inefficient way of comparing population distributions (9).

A simple way to test for changes in the mean and SD of BMI is to convert them to $z$ scores, using a convenient LMS-based reference. The mean and SD ought to be 0 and 1, respectively, so that if either is materially increased, then it provides evidence of a trend to increasing obesity. Bundred et al. (30) demonstrated clear trends to increasing BMI over 10 years in 3- to 4-year-old British children in this way: they converted BMI to $z$ scores using the British 1990 reference, and the mean $z$ score increased by 0.4 units over the 10 years, whereas the $z$ score for height did not change. Unfortunately, the paper did not report the corresponding SDs over time.

CONCLUSIONS

BMI is a simple though imperfect measure of overweight in childhood. Reference standards for BMI are important both for clinical and population use in the study of
child obesity, to monitor the prevalence and incidence of overweight and obesity. However, there are serious difficulties with nationally based references for making international comparisons, and a different approach is needed to provide a politically acceptable definition.

REFERENCES


DISCUSSION

Dr. Dietz: I have a comment and a question. My comment is that I did not see persistence of obesity into adulthood among your morbidity criteria. I think there are now sufficient longitudinal studies to allow that to be added as an additional criterion. My question has to do with how one might develop an agenda for defining obesity among Asian children. As you know, in adults the cutoff point appears to be lower, for reasons that are not clear. There is substantial morbidity at lower adult BMIs than is apparently the case in Western societies, which would imply that the cutoff points in, for example, Chinese and Japanese children might well be lower. How would you define the research agenda to explore that issue?

Dr. Cole: Once you accept that the risk in Asian children is greater, then implicitly you are saying that you have got to use a risk-based cutoff. If you are to use a risk-based cutoff, this leads us back to the problem that we don’t have a very clear outcome measure that we can use. I think in a sense this just underlines the need for us to try and move forward in this area to clarify our ideas about how to obtain a risk-based cutoff. The other thing we need to remember is that the concept of a “cutoff” is a great oversimplification. We know that the underlying continuum is just as important and that it does not really matter whether you are just below the cutoff point or just above it. This is the danger of cutoffs, and it goes against my statistical philosophy to try to pretend that you have a distribution that is present or absent when we know that that is absolutely not the case. I could probably live with a cutoff that did not have any implications for risk, while accepting that if it were applied to Asian children it would have a different meaning from that in European children—in other words, a given value was likely to be more serious in terms of outcome.

Dr. Uasy: When you approach the change with age, do you take different rates of maturation into account? Children who mature earlier will have a higher BMI during the pubertal period.

Dr. Cole: This is a clinical rather than a population issue. In a population, what matters is the average timing of maturation. If they have early maturation, they will show higher rates of overweight at that time in relation to the IOTF cutoff, but once they have passed through puberty the figures will fall back to where they were before. I get nervous when I’m asked this sort of question. It suggests that these cutoff points are wonderfully detailed instruments, but in fact they are about as blunt as you could imagine. They are not designed to give a good answer in individual clinical situations, and maturation is an issue in point. I am sure one could think of lots of others. However, in terms of population studies, and taken broadly over a range of different ages, I think one can get around those difficulties.

Dr. Koletzko: I think it is a big step forward that you have created a global standard for obesity prevalence and changes in incidence. You started your introduction by saying that there is a broad consensus approving the use of the BMI as a standard measurement in children, but now you appear to be questioning that. I wonder whether it is premature to apply the BMI as the standard clinical measure, based on the limited scientific evidence that we have and the various practical limitations. For pediatricians in clinical practice, it is much more inconvenient to calculate the BMI and to look at age- and sex-specific reference values than to use the conventional weight-for-height centiles. I haven’t seen any evidence that the use of the BMI in clinical practice is better at identifying children at risk or that it improves the outcome. You indicated—and
I could not agree more—that what we really need is a risk-based approach. Dr. Dietz has already posed the question of standards for a risk-based approach. For example, there are now definite indications that children of low birth weight who develop a high BMI have a much greater risk of metabolic and clinical consequences than infants with the same BMI who had a higher birth weight.

Dr. Cole: Your first point was about the consensus over whether the BMI is a good measure to use in children in clinical practice. I'm not a clinician, but I'm not clear that there is any real alternative to using weight-for-height, and you come back to that at the end. Pediatricians have been working with weight-for-height as a concept since 1966—that is, they have at least 40 years and possibly 50 years of familiarity. However, we know that weight-for-height is not a very good index under certain circumstances. For example, in infancy and adolescence it does not actually work. You'll know that the NCHS weight-for-height chart stopped before you got to adolescence, so for the subject of our meeting here, which mentions adolescence, the NCHS weight-for-height reference would not have been much use. So weight-for-height, which clinicians are very happy to use, didn't actually do the job, while BMI does do the job. So my conclusion is that there is probably clinical support for using it. I accept that it's slightly more difficult to work with because you have to get out your calculator to work it out.

Dr. Koletzko: I'm a friend of the BMI myself. I'm just asking, at this time of evidence-based medicine, whether we are right to tell the world of clinical pediatrics to use the BMI if there is no evidence for any benefit of BMI over weight-for-height. You mentioned adolescence and with the dataset you have you could easily construct a weight-for-height centile for adolescents as well. So the fact that the weight-for-height relation changes in adolescence is not a good argument for using the BMI in my view. I have a question about the application of your standards to German children. When we apply your standards in Germany, then in comparison with our traditional weight-for-height standards it appears that your cutoff point identifies a much lower proportion of children as obese. Do you have an explanation for that?

Dr. Cole: It depends what age you are looking at. As I said, NCHS tried to recognize this difficulty—that if you use weight-for-height in the simple sense in adolescence it really doesn't mean anything. If you try to compare BMI and weight-for-height in adolescence (was this the age range you were referring to?), you will get answers that look very different, and you will have to take it from me—although you may not accept it—that BMI is the more correct of the two answers. If the difference you are referring to are in earlier childhood, then the explanation may simply be that the cutoff points are different, so you will get different apparent rates of obesity. This just underlines the fact that the definition is entirely arbitrary. However, the advantage of the IOTF standard is that it will give the same answer for children of the same size in different countries.

Dr. Birch: I believe these reference data are based on cross-sectional data. It concerns me, from my experience in other areas, that you can get yourself into trouble by drawing inferences about change over time when using cross-sectional data, particularly in an area such as this where there are rapid changes in prevalence, and probably time-of-testing effects. With this in mind, can we really say that crossing centiles is a reasonable way of looking at things, when we are looking at an individual child but comparing that child with cross-sectional group data?

Dr. Cole: Your point is well made. I think it is important to remind people that all charts, from way back, have been based on cross-sectional data, both for weight and height, and yet the concern about centile crossing on those charts tends not to get discussed. I agree you do require longitudinal data to judge the amounts and the significance of centile crossing. We know that the average child is crossing the centiles upward, that charts are getting out of date day by day. This leads on to another debate about whether we should make our charts so that they represent
children as they are now, or as they used to be in a halcyon age when everybody was the right fatness. There isn’t a simple answer to that, but you do need longitudinal data to assess centile crossing, I agree, and such data are rather sparse.

**Dr. Sivastava:** You mentioned cutoff points based on risk, but risks may vary from country to country. How do we sort this out and relate risk clearly to obesity?

**Dr. Cole:** It is a very difficult problem, and it just highlights the difficulty of using any sort of risk-based cutoff. The association between childhood BMI and later outcome is so diffuse and so varied from country to country that it’s difficult to pin down the particular outcomes that one wants to focus on, or to come up with such a clear relationship that you are prepared to draw a cutoff on a chart as a result of it. This is why people are not keen to use a risk-based cutoff. There is too much uncertainty.

**Dr. Bar-Or:** As physiologists, we pride ourselves in our ability to assess the percentage of body fat much more precisely than you can by using BMI or skinfold techniques. Even so, when we arrive at a value for percentage of body fat, we don’t really know what it means as far as cutoff points are concerned. In your various datasets, have you encountered subsets that could allow you to correlate the BMI with a more precise measurement of body fat, so that we can make sense of our own data?

**Dr. Cole:** I think you put your finger on it when you said that we don’t know what *percent body fat* means. The bottom line is that we want to come up with some anthropometric indicator of outcome. If that anthropometric indicator happens to be a refined estimate of percent body fat, then so be it. But it might not actually relate to percent body fat particularly well at all. The important thing is that it predicts outcome. So there is a lot of work to do. What we need is an indicator of what is going to happen in the future, so the percent body fat issue is almost a red herring. It might be the solution, but we actually want something that is going to be useful in the long term, and the big difficulty is that we cannot get long-term data other than by hanging around a long time.

**Dr. Usa: I’d like to make a point that the WHO standard was a “gueestimate” for risk for a Caucasian population. I don’t think it is a valid end point for children until we know the risks associated with given BMIs. In fact, I would for now avoid calling it a “standard”; it would be preferable to call it a reference. I don’t think it has been validated, either in terms of body composition or in terms of risk. Both for populations and for individuals we need a standard that is related to risk and to body fat.**

**Dr. Cole:** I agree with that.