Comparison of weight and height relations in boys from 4 countries

MF Franklin

ABSTRACT Height and weight data from children in the United States, United Kingdom, Japan, and Singapore were analyzed to investigate differences in growth between the groups of children. An investigation into the adjusted weight indexes of the form index = weight/height^p for differing powers of p (Benn index) showed that the power of p required to produce a correlation of zero between the index and height varied with age. For the United States, Japan, and Singapore the p value was just below 3.0 for children aged 6 y, increased to ~3.5 for children aged ~10 y, and decreased to ~2.0 at age ~18 y. A consequence of the p value being mostly > 2.0 is that BMI (wt/ht^2) tends to be greater for tall children than for short children. The US data (from the second National Health and Nutrition Examination Survey) also contained information on skinfold thickness. Relating skinfold thickness to indexes of the same form for height and weight suggested that the best relation was achieved with p values of ~2.0 except for children aged 12–16 y, for whom the optimal values for p were higher. The highest value, 2.9, was achieved at ages 12–13 y. Overall, the use of BMI as an indicator of adiposity appears acceptable for children aged 6–7 and 17–18 y. However, BMI should be used with caution when assessing children aged 8–16 y. Am J Clin Nutr 1999;70(suppl):157S–62S.

KEY WORDS BMI, weight, height, obesity, children, Benn index, United States, United Kingdom, Japan, Singapore

INTRODUCTION Epidemiologic studies require a simple indicator of adiposity. Such a measure should be applicable under a wide range of conditions, including those in which no advanced technical equipment is available or the staff has received little training. Unfortunately, many direct measures, including skinfold thickness, are subject to interpretation and may differ markedly between observers and studies. It is therefore risky to compare results of surveys unless standard procedures were used throughout. Lack of consistency in the measuring process not only affects comparisons between surveys conducted in different places at the same time but also comparisons between studies carried out in the same population at different times.

The simplest indicators of adiposity are those based on weight, particularly when each individual’s weight has been suitably adjusted for sex, age, and height. Because weight, height, age, and sex are relatively easy to obtain and not greatly affected by observer bias, they provide a relatively stable, if somewhat indirect, assessment of adiposity. Although the links between anthropometric indicators and adiposity, and more particularly various health risks, are somewhat loose, they have been accepted as informative at the population level, though not necessarily at the individual level. Studies in children have used indicators based on the tables of weight and height for age and weight-for-height tables produced by the US National Center for Health Statistics (NCHS) and the World Health Organization (WHO) (1). For adults, assessments have been based on body mass index [BMI (ht/wt^2)]. In this exercise we require a measure applicable to a wide range of child populations and with a similar meaning in each. The measure should apply to children aged 6–18 y but should be compatible with measures for adults aged ≥19 y.

Weight is affected by sex, age, and height; height is also affected by sex and age. The relations between height and weight may be affected by numerous other factors, including ethnic origin and social class. In adults, however, these relations are simpler because adult height does not change. Further simplifications arise because 1) although adults tend to gain weight with age, methods for defining obesity tend not to vary with age; and 2) the effect of sex on weight is viewed as a consequence of its effect on height. Thus, weight is adjusted for height only. Typically this adjustment takes the following form

adjusted weight = weight/height^p

where p is a value between 1 and 3. An index of this form is commonly called a Benn index (2); in the special case that p = 2.0 it is called the BMI. In an ideal world of individuals of the same shape and body density the p value would be close to 3.0, i.e., weight would be adjusted for volume. In a population in which height and weight are not related, the p value would be close to zero. The “best” value for p is often estimated by fitting a linear relation of the form

log(weight) = p × log(height) + c

where c is a constant.

1 Undertaken on behalf of the International Obesity Task Force working group on childhood obesity.

2 Reprints not available. Address correspondence to MF Franklin, Callater Lodge, Glenshee Road, Braemar, Aberdeenshire, Scotland AB35 5YQ, United Kingdom. E-mail: mike@hotel-braemar.co.uk.
A problem with indexes determined in this way is that the \( p \) values obtained vary both within and between studies, making comparisons between groups difficult. For adults, the problem is to some extent overcome by standardizing the value \( p = 2 \), i.e., BMI, which, though not optimal, is nearly optimal in most studies. The use of BMI is widely accepted, so that BMI values > 25 and 30 are commonly deemed to indicate individuals who are mildly obese and obese, respectively, for most ethnic groups, irrespective of sex, age, height, and social class.

The wide acceptability of BMI for adults does not necessarily extend to children because of the effect of age on the height in children. Also, even for children of similar age the estimated values for \( p \) may not be close to 2.0. The use of BMI to monitor weight in child studies is therefore questionable and even if used it would probably require more complex treatment than in adult studies. The use of BMI for children is advocated by Cole (3), who produced reference curves of BMI with age for British children (4). More generally, for children, the expected weight may be modeled by a function that increases with both age and height, whereas the expected height may be modeled by a function that increases with age. If, however, the weight-to-height relation does not change with age it is possible to simplify the model by relating weight to age and height separately. To some extent, such simplifications are built into the NCHS tables (1), which provide both weight-for-height and weight-for-age standards. However, because the compilers of these tables were cautious of the effect that puberty would have on validity, the maximum heights allowed in the weight-for-height tables correspond to the median heights of 10-y-old girls and 11-y-old boys. Beyond these ages the likelihood of puberty increases and with it a possible change in the weight-to-height relation. The replacement of NCHS standards by other methods of adjusting weight for height is unlikely to remove the influence of puberty.

The functions relating weight, height, and age tend to be smooth but do not have a simple parametric form. Therefore, in most studies on children, the effect of age on height or weight is generally estimated through reference to charts showing the performance of a standard population. Individuals are then assessed against this population through use of SD scores \((z\) scores) or population percentiles. There has been much debate about how the standard population should be chosen, e.g., from NCHS standards based on data from US children, from local national standards, or from standards determined from data from many countries. There has also been debate on how best to develop the charts, given that even for the standard population not all children were assessed.

In this article, some analyses of 4 data sets from the United States, United Kingdom, Japan, and Singapore are presented. The objective was to provide insights that may aid in the selection of the best methods of forming standards, specifically to 1) seek similarities and differences between the 4 groups; 2) investigate the relations between weight, height, and age; 3) investigate which values of \( p \) are the most appropriate for adjusting weight; and 4) determine whether BMI \((p = 2)\) is as valid for children as for adults.

**METHODS**

Anthropometric data for children from the United States, United Kingdom, Japan, and Singapore were obtained from various sources as described by Guillaume (5). All 4 data sets were drawn from communities that can be regarded as adequately nourished, and all were treated as coming from cross-sectional studies. The data sets were not gathered as part of a coordinated study, and the data from United States and United Kingdom were collected well before those from Japan and Singapore. Except for Japan, where the oldest children reported were age 14 y, the analyses reported here relate to boys aged 6–18 y. For the United States and Singapore, results were further restricted to white and Chinese children, respectively. The number of boys in each year category ranged from 127 to 1017 (Table 1).

**Relations between weight, height, and age**

For each group, the boys were categorized by age group, either in 6-mo or 1-y intervals, and the mean height and weight were determined for each category. These means were then plotted against age and the resulting growth curves were compared to expose differences and similarities between the growth of the 4 groups.

In addition to these height-to-age and weight-to-age relations, the mean weights and heights were transformed by logarithms and plotted against each other. The graphs for the 4 groups were again compared; such graphs illustrate the interage relations between weight and height. The slopes of the lines yield an estimate of \( p \) that may be compared with the \( p \) values obtained for each age class. They may also be compared with the weight-for-height lines produced by the NCHS. (This latter comparison, however, should not be interpreted too strictly because the NCHS relations were derived from boys of similar heights but different ages, whereas the current means were derived from boys of similar age but different heights.)

**Estimating the \( p \) value for each age class**

For each group of boys, attention was given to individual age classes. For each class the relation between weight and height may be modeled through the simple linear regression function of the form shown by Equation 2 above to yield an estimate of the value for \( p \). The values for \( p \) were then plotted against the mean age for the class and the 4 curves were compared.

**Percentile curves**

The percentile plots presented in this article were produced in various ways. Those for US data were derived from the work of Must et al (6), who determined the percentiles for each age class and then, assuming that percentiles varied with age in a smooth manner, replaced the observed values with smoothed values. Percentile plots for UK data were derived from the curves of Cole (7) using a program written by Cole et al (4). These percentiles were produced essentially as follows. For each age the authors determined a power transformation that best converted the data to normality. For the transformed variable they determined the mean and SD. They then assumed that the power, mean, and SD varied.

**TABLE 1**

<table>
<thead>
<tr>
<th>Country</th>
<th>Maximum age</th>
<th>Number of boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>14</td>
<td>3458 (384)</td>
</tr>
<tr>
<td>Singapore</td>
<td>18</td>
<td>13218 (1017)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>18</td>
<td>4149 (230)</td>
</tr>
<tr>
<td>United States</td>
<td>18</td>
<td>1646 (127)</td>
</tr>
</tbody>
</table>

\(^1\)Data from reference 5. Average number of boys per year of age in parentheses.
with age in a smooth manner and replaced the observed values with smoothed values. Finally, these smoothed values were used to estimate the percentiles on the transformed scale, which were then back-transformed to give percentiles in the original scale. For the Japanese and Singaporean data, percentiles were obtained directly from the data and no smoothing was performed.

The US data were used to determine whether height could have an effect on classifying for obesity as follows. First, each boy was classified as exceeding or not exceeding the 85th percentile as calculated by Must et al (6) on the basis of the same data set. Second, each boy was classified as belonging to the first, second, or third height tertile. If height has little or no effect then the proportions of boys exceeding the 85th percentile should be roughly equal for all height categories and age groups. The actual proportions observed were plotted against age to display discrepancies and the effect of age, if any, on these discrepancies.

Relating skinfold thickness to weight and height

US data from the second National Health and Nutrition Examination Survey (NHANES II) (6) and the National Health Examination Survey (NHES) (8) were used to seek values that would give adjusted weights of the form $wt/ht^p$ that would relate most closely to skinfold thicknesses. The values of $p$ were determined as follows. When some function of skinfold thickness, $f(skf)$, is built into a linear model with $\log(wt)$ and $\log(ht)$, then it is possible by using standard regression methods to estimate the power of $p$ that leads to the best relation between adjusted weight and skinfold thickness. However, there are 2 sensible ways of formulating such a model:

$$f(skf) = a + b \times [\log(wt) - p \times \log(ht)]$$ (3)

$$\log(wt) = -a \times b^{-1} + b^{-1} \times f(skf) + p \times \log(ht)$$ (4)

where $a$ and $b$ are coefficients to be estimated.

These models lead to different estimates for $p$ because different assumptions are made about the errors associated with them. Consideration of the effects of these error assumptions on the estimates suggests that the value obtained from Equation 3 will underestimate $p$, whereas that from Equation 4 will tend to overestimate it. For each age class the value for $p$ was estimated by using both methods in data from the US boys, and the estimates from Equation 4 were indeed generally larger than those from Equation 3. In this article the 2 estimates were averaged by forming a geometric mean and plotting this mean against age.

RESULTS

Relations between weight, height, and age

Up to the age of 15 y the UK and Singaporean boys were of similar stature and both were slightly shorter than the US and Japanese boys (Figure 1A). After age 15 y the Singaporean boys appeared not to gain height whereas the UK boys did; by age 18 y they were similar in height to US boys. The Japanese data set contained no data for boys aged ≥15 y. The weight curves show a similar pattern to the height curves, but the UK boys weighed less than the US boys at all ages (Figure 1B).

In Figure 2 the mean weights and heights shown in Figures 1A and 1B are presented on a log-log scale. Here the curves are remarkable for their similarity, the 4 lines being difficult to distinguish. The UK children tended to weigh slightly less for a
time with a slope equal to 2.58 (data not shown). For a boy with a height of 130 cm, the NCHS tables suggest a median weight of 26.8 kg compared with an estimated median weight of 27.1 kg of the US boys in the analysis presented here.

**Estimating the p value for each age class**

The values for the slope p (estimated from Equation 2) in each age class for the 4 data sets are shown in Figure 3. Clearly, each curve shows a pattern and, by and large, the pattern was similar in all the data sets. For the US, Japanese, and Singaporean boys, the p value was <2.8 at age 6 y, increased to <3.5 at age 9–10 y, and steadily decreased to <2.0 at age 16 y. The p value was mostly ≥2.8 for all ages up to 13 y. The p value for the UK boys followed a similar but flatter trend, with the p starting at <2.3, increasing to <2.6 before decreasing once more to <2.0. Possibly the most surprising feature of the curves was the existence of slopes >3.0, suggesting that even volume does not provide a full explanation for weight differences. The slopes of <2.0 for boys aged ≥16 y agreed with the use of a p value of 2 in BMI for adults.

The consistency with which p > 2.0 suggests that BMI will underestimate the effect of height on weight and that it will tend to increase with height. This last point is readily illustrated because the slope of the regression line for relating log(BMI) to log(ht) is p – 2, which exceeds zero whenever p is >2.0. The median BMI will therefore increase with age, and taller children will tend to have larger BMIs than will shorter children.

**Percentile curves**

The 85th and 95th percentiles for BMI plotted against age increased in boys from each of the 4 countries (Figure 4). The percentiles were similar for all groups of boys up to age ≈9 y. The BMI values, particularly the 95th percentile, for the US boys increased faster with age than those for the UK boys. The graph
for the Singaporean boys suggests a curvilinear change not visible in the other data sets.

The proportions of boys, from the US data, above the 85th percentile of BMI in 3 height tertiles as defined by Must (6) are shown in Figure 5. The figure illustrates clearly that for younger children the proportion exceeding the percentile was much greater in the tallest class than in the shortest class. For the older boys there was little difference between the proportions exceeding the 85th percentile.

Relating skinfold thickness to weight and height

The $p$ values that yielded the best relation between the Benn index and skinfold thickness with the NHANES II data (6) are shown in Figure 6. Also shown in this figure is the equivalent graph obtained from the data of another US survey, the NHES (8). Estimates for $p$ were generally $>2.0$ for most ages but between 12 and 16 y of age the $p$ value was $>2.0$.

DISCUSSION

One of the dangers of basing standard growth curves on a heterogeneous population is that the spread of observations about the median line, as measured by the SD, could be very large. Therefore, although the family of growth curves may be representative, they could prove insensitive as a measurement tool. The 4 groups of boys analyzed in this study had similar growth patterns, but UK and Singaporean boys were slightly shorter than the US and Japanese boys. Differences in weight between the groups were largely caused by differences in height. The similarities suggest that it should be possible to combine the results of several countries to develop standard growth curves of wide applicability and adequate sensitivity. The similarity between the growth patterns of the US, Japanese, and Singaporean boys was quite striking in that it even extended to similar patterns for the way that the index power $p$ varied with age. Further analysis to determine why the UK boys yielded lower powers of $p$ could prove informative.

The analyses presented in this article underscore the need to be cautious when BMI is used as an indicator of adiposity in children. BMI adjusts weight by a factor equal to height$^2$, but the analyses generally indicated that the use of a power $p$ for height where $p > 2.0$ would give a better measure for children ages $\leq 15$ y. These higher values of $p$ occurred in the within-year analyses, the between-year analyses, the original NCHS weight-for-height tables, and the relations with skinfold thicknesses. Also, the proportion of tall children being classified as overweight (according to the 85th percentile) is greater than that of short children, especially in younger children. The use of BMI, however, appears to be satisfactory for adolescents aged $\geq 16$ and links well with the use of BMI for adults. Moreover, although the use of a fixed power $p > 2$ removes some of the weaknesses associated with BMI in children, it does not remove all of them, suggesting that a different index be used for adults and children. Use of a variable power that is $>2.0$ for younger children but decreases to 2.0 for children aged 16 and above, removes yet more of the weaknesses but again at a cost of extra complexity with its potential for errors. With traditional ways of assessing adiposity through weight and height (eg, pencil, paper, and reference tables) there is probably little advantage to basing an index on a power $p$ other than 2.0. However, with increasing availability of low-cost computers there could be a case for basing software around alternative values of $p$, perhaps even allowing it to vary (smoothly) with age.

Three different methods of determining percentiles were used in this study. For the US and UK data sets the percentiles used had been published elsewhere. For the Japanese and Singaporean data we calculated the percentiles directly from the data for each age class and no attempt was made to smooth the percentile-age curves. Each of the 3 methods has advantages and disadvantages. The Must et al (6) procedure is direct, but for small data sets the between-year analyses, the original NCHS weight-for-height tables, and the relations with skinfold thicknesses. Also, the proportion of tall children being classified as overweight (according to the 85th percentile) is greater than that of short children, especially in younger children. The use of BMI, however, appears to be satisfactory for adolescents aged $\geq 16$ and links well with the use of BMI for adults. Moreover, although the use of a fixed power $p > 2$ removes some of the weaknesses associated with BMI in children, it does not remove all of them, suggesting that a different index be used for adults and children. Use of a variable power that is $>2.0$ for younger children but decreases to 2.0 for children aged 16 and above, removes yet more of the weaknesses but again at a cost of extra complexity with its potential for errors. With traditional ways of assessing adiposity through weight and height (eg, pencil, paper, and reference tables) there is probably little advantage to basing an index on a power $p$ other than 2.0. However, with increasing availability of low-cost computers there could be a case for basing software around alternative values of $p$, perhaps even allowing it to vary (smoothly) with age.

Three different methods of determining percentiles were used in this study. For the US and UK data sets the percentiles used had been published elsewhere. For the Japanese and Singaporean data we calculated the percentiles directly from the data for each age class and no attempt was made to smooth the percentile-age curves. Each of the 3 methods has advantages and disadvantages. The Must et al (6) procedure is direct, but for small data sets the extreme percentiles are poorly estimated, and when plotted against age the points on the line can appear erratic, making it difficult to judge how much smoothing would be appropriate. Furthermore, the curve for each percentile is smoothed independently of the others so that the smoothed percentile curves can (somewhat illogically) cross. This crossing is evident in some of the lines Must et al (6) present for black children. The method of Cole (7) avoids this problem, but the process is somewhat indi-
The effect of a “wrong” choice for estimates of both mates obtained are highly correlated so that different values for where the more general power transformation over, the transformation suggested by Cole (7) is a special case of the more general power transformation

\[ y = (x + c)^{b} - 1 \]

(5)

\[ y = \log(x + c) \]

(6)

where \( x \) is the original variable, \( y \) is the transformed variable, and \( b \) and \( c \) are variables chosen to make the distribution of \( y \) as close to normal as possible.

Cole lets \( b \) vary but sets \( c \) to zero. Other authors, eg, Rosenbaum (9) set \( b \) to zero and use the logarithmic transformation. This transformation sometimes has the advantage of stabilizing variance. However, it can be shown that when the maximum likelihood procedure is used to obtain estimates for \( b \) and \( c \), the estimates obtained are highly correlated so that different values for both \( b \) and \( c \) can apparently lead to equally good transformations. The effect of a “wrong” choice for estimates of \( b \) and \( c \) is most likely to be apparent for extreme percentiles, ie, the 95th, which is used in some obesity studies. It is essential, therefore, that the use of the transformation method be accompanied by detailed checks on how well it models these extreme percentiles.

Comparisons of children on the basis of BMI is clearly superior to comparisons on the basis of their unadjusted weights. However, for children aged \( \leq 15 \) y the adjustment is only partial so that BMI is positively correlated with height. Therefore, to interpret a child’s BMI, it is also necessary to know the child’s age, height, and probably sex. The need for these data contrasts with adults in which such extra information is often deemed unnecessary. The dangers of using BMI may be illustrated by assuming that A and B are 2 populations with the same weight and height relation, but that population B is taller than population A. (A and B could be 2 different groups of children or the same group assessed on 2 occasions.) Assessment of the 2 populations on the basis of BMI will indicate that B has a higher prevalence of obese children than A because taller children tend to have higher BMIs and hence are more likely to be classified as obese. Thus, it might be concluded wrongly that B is in greater need of intervention. Thus, the BMI should not be used without additional information with which to classify individuals.

BMI is not obviously superior to the weight-for-height criterion built into the NCHS standards. The constraints placed on these latter tables by the onset of puberty are likely to extend to any form of weight-for-height index, including BMI. However, the relations between \( \log(\text{ht}) \) and \( \log(\text{wt}) \) were essentially straight for all 4 groups over the age range of 6–18 y and do not suggest that a marked change occurs with puberty. This suggests that 1) the height restrictions on the NCHS weight-for-height tables are possibly somewhat conservative and 2) if BMI is interpreted with respect to age and height, and the onset of puberty is similar for both study group and the standard population, BMI should prove useful for assessing the nutritional status of children if used properly.

The International Obesity Task Force childhood obesity working group thanks U Rajan for providing the Singaporean data set, T Cole for the UK data, A Must for the US data, and Y Matsuzawa for the Japanese data set.

REFERENCES