Curve smoothing and transformations in the development of growth curves\textsuperscript{1,2}

Katherine M Flegal

ABSTRACT Growth charts such as those published by the National Center for Health Statistics (NCHS) consist of a set of smoothed percentile curves showing the distribution of different aspects of body size for infants, children, and adolescents. The original NCHS growth charts are currently being revised to incorporate additional national data, to include growth curves for body mass index (BMI, in kg/m\(^2\)), to reduce or eliminate discontinuities, and to use new and improved methods of smoothing. Methods of curve smoothing in the development of growth curves are briefly discussed in the context of this revision by using examples based on the provisional data set for revision of the growth curves. Among the factors that may affect the construction of the revised growth curves are sample sizes, sampling weights, and secular trends. The use of BMI or other weight-height indexes in growth curves presents some new issues because these transformations of weight and height data do not increase monotonically with age. Some of the advantages and disadvantages of several different approaches to creating smoothed percentiles and standardized scores are discussed briefly. These effects need to be considered in the broader context of constructing growth curves. No single method of developing smoothed curves is clearly the best for all purposes, and estimates of overweight and underweight may be sensitive to the method chosen. Alternative approaches to constructing smoothed curves by using weight-height functions other than BMI might warrant further exploration. Am J Clin Nutr 1999;70(suppl):163S–5S.

KEY WORDS Anthropometry, body weight, body height, body mass index, growth, growth charts, growth curves, growth percentiles, health surveys, curve smoothing, children, adolescents, ponderal index

INTRODUCTION Growth charts such as those published by the National Center for Health Statistics (NCHS) in 1977 consist of a set of smoothed percentile curves showing the distribution of different aspects of body size for infants, children, and adolescents (1). The original NCHS growth charts are currently being revised to 1) incorporate additional national data and possibly also information on the distribution of birth weights in the United States, 2) include growth curves for body mass index (BMI; in kg/m\(^2\)), 3) reduce or eliminate discontinuities, and 4) use new and improved methods of smoothing (2). Methods of curve smoothing in the development of growth curves will be discussed in the context of this revision. The examples presented here are based on the provisional data set for revision of the growth curves for boys aged 2–19 y.

Three factors that may affect the construction and smoothing of the revised growth curves are sample sizes, sampling weights, and secular trends. For the revised charts, the present plan is that US data from 5 nationally representative cross-sectional surveys (3) conducted between 1963 and 1994 will be combined and the sampling weights from each survey will be used without reweighting of the data. For older children, there are \textless 100 or fewer children of each sex per month of age. This sample size is adequate but not generous. In this situation, it is desirable to use methods of estimation and smoothing that, to the extent possible, incorporate information from neighboring ages to include more data. Sampling weights are used to adjust for differential probabilities of selection. The sampling weights make estimation and smoothing more complex and add to the variability of the data.

Secular trends in childhood weight also complicate the picture. Data from the most recent US national survey, the third National Health and Nutrition Examination Survey (NHANES III), showed that the prevalence of overweight among children increased during the time period covered by the NHANES surveys (3). The question of whether it is appropriate to exclude the heavier NHANES III children from the revised growth reference data has been a matter of some debate.

WEIGHT, HEIGHT, AND FUNCTIONS OF WEIGHT AND HEIGHT

The unsmoothed (empirical) percentiles of height and weight for age from the provisional working data set for revision of the National Center for Health Statistics/Centers for Disease Control growth charts are shown in Figure 1. The unsmoothed height percentiles are fairly regular and symmetric and are close to a normal

\textsuperscript{1}From the National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD.
\textsuperscript{2}Reprints not available. Address correspondence to KM Flegal, National Center for Health Statistics, Centers for Disease Control and Prevention, 6525 Belcrest Road, Room 900, Hyattsville, MD 20782. E-mail: kmf2@cdc.gov.
distribution; they also generally increase monotonically with age and show relatively little variability otherwise. The unsmoothed weight percentiles also tend to increase monotonically with age but show a more skewed distribution and more variability.

BMI is a widely used function of weight and height that may itself be considered a type of transformation of the weight data. Ponderal index (PI; height/weight^{1/3}), an index suggested originally by Sheldon et al (4), is used here to illustrate the behavior of a different weight-height function (5). Because PI increases as weight decreases, \((100 - \text{PI})\) is used instead to provide a weight-height index that varies with weight in the same direction as does BMI.

The raw percentiles of BMI and \((100 - \text{PI})\) for age are shown in Figure 2. Percentiles of both indexes are skewed and highly variable, particularly at the upper percentiles, and do not increase monotonically with age. Percentiles of BMI tend to dip slightly with increasing age from 2 to \(<6\) y and then rise steadily with age after \(<6\) y. In contrast, percentiles of \(100 - \text{PI}\) tend to decline linearly with increasing age up to \(<6\) y and then remain relatively constant over the rest of the age range. There might be some advantage within a given age range to using a weight-height index, such as PI, that varies little with age within the range.

METHODS OF SMOOTHING

Smoothed reference data from growth charts are often presented as either percentiles or \(z\) scores (standard deviation scores) (6). In the United States, percentiles are the most commonly used method for clinical monitoring of individual growth. For height and weight, percentiles increase monotonically with age; this property needs to be preserved in the smoothing. In addition, curves should not be oversmoothed (losing too much variability) or undersmoothed (preserving too much variability). An example of the possible effects of oversmoothing is provided by the published smoothed BMI percentiles from the first National Health and Nutrition Examination Survey (NHANES I) (7). For girls aged 6–17 y, the published smoothed 85th percentile points lie below the empirical percentile points almost everywhere along the curve and are lower on average by 0.6 BMI units. The slight
oversmoothing has a surprisingly large effect on the estimated prevalence of overweight based on the 85th percentile of BMI. When the published smoothed 85th percentiles are used, the estimated prevalence of overweight is higher by 3 percentage points for girls aged 6–17 y and by almost 5 percentage points for girls aged 6–10 y than when the empirical 85th percentiles are used.

One approach to generating smoothed percentiles is to first calculate empirical percentiles and then smooth those percentiles. Examples of parametric or nonparametric methods that can be used for smoothing include cubic splines, kernel regression, locally weighted regression, and running medians. After comparisons of these methods were made, including assessment of root mean square error, locally weighted regression was selected and is currently being applied to the provisional data set for growth chart revisions (8). This method provides smoothed percentiles but not \( z \) scores. It requires few assumptions and the results tend to be close to the empirical data. Sampling weights are readily accommodated.

A second approach, exemplified by the lambda-mu-sigma (LMS) method, is to model the data, smooth the model parameters, and then estimate smoothed percentiles from the model parameters (9). The LMS method models the entire distribution taking into account degree of skewness (\( L \)), central tendency (\( M \)), and dispersion (\( S \)). The \( L \), \( M \), and \( S \) parameters are estimated and then smoothed using any of a variety of methods. Any desired percentile or \( z \) score can then be calculated from the smoothed \( L \), \( M \), and \( S \) parameters. This method requires more assumptions and it is less obvious how sampling weights should be incorporated. Its advantages are that it permits calculation of \( z \) scores as well as percentiles and allows calculation of any desired percentile.

Accurate estimation of percentiles from the LMS method relies on the assumption that after transformation and smoothing, the variables of interest are normally distributed. To investigate the validity of this assumption for the revised growth chart data, the distribution of \( z \) scores calculated using smoothed values of \( L \), \( M \), and \( S \) was compared with a normal distribution. For weight and BMI there was a deviation of almost half a standard deviation at the upper end of the distribution for one age group, as well as an apparent slight systematic departure from normality, suggesting a possible need for some further transformation.

Departure from normality in the tails can affect estimates of underweight or overweight based on BMI percentiles. For boys aged 2–19 y, estimates of overweight and underweight based on modeled (LMS) 90th and 10th percentiles of BMI were in several subgroups ≥2 percentage points lower than estimates based on the smoothed or unsmoothed empirical percentiles or the raw data. For example, of boys aged 10–13 y, 10% fell below the empirical 10th percentile but only 7.5% fell below the modeled (LMS) 10th percentile.

**CONCLUSIONS**

These effects must be considered in the broader context of constructing growth curves, including decisions on inclusion or exclusion of data points and methods of estimation. Other aspects besides transformation and smoothing procedures may well have effects of similar magnitude on estimates of overweight or underweight. A variety of methods is available for developing smoothed curves. There is no single method that is clearly the best in all situations and for all purposes. It is likely that a number of methods work similarly well and produce fairly similar results. However, estimates of overweight and underweight may be sensitive to the methods chosen. Alternative approaches to constructing smoothed curves using weight-height functions other than BMI may warrant further exploration.

**REFERENCES**