Modified z-scores in the CDC growth charts

Data quality assessment on anthropometry data:

In the analysis of childhood body size measures, it is often necessary to identify very extreme values. These extreme values are often considered to be biologically implausible values (BIVs). These extreme values may represent errors in data entry or measurement, but it should be realized that they may also simply represent correctly recorded values that are very high or low [1,2].

Z-scores (or standard deviation (SD) scores) are widely used in anthropometry to quantify a measurement's distance is from the mean. For example, a measurement that is 1.645 SDs from the mean (z-score of 1.645) would be at the 95th percentile of a normal (0,1) distribution; 95% of the distribution would be less than this measurement.

The following text describes the calculation of z-scores and 'modified z-scores' in the CDC growth charts for BMI values. Similar procedures were used to derive the z-scores and modified z-scores for weight and height.

CDC growth charts and the LMS method

The CDC growth charts contain 10 smoothed percentiles (between the 3rd and 97th) of BMI for 24 to 240 month-old children and adolescents [3]. These smoothed estimates were subsequently used to derive lambda (L, the power transformation to achieve normality), mu (M, mean or median) and sigma (S, coefficient of variation) parameters for Cole's LMS method [4,5]. Estimates of these 3 parameters allow the BMI of any children to be expressed as a z-score (BMIz) and percentile relative to children of the same sex and age in the CDC growth charts.

However, it should be noted that the approach used in the CDC growth charts to estimate these 3 parameters differs from that proposed by Cole [6,7]. In the CDC method, the LMS parameters for BMI were derived from the equations for smoothed percentiles, whereas the original method derived these parameters from the entire population.

To calculate the z-score for any BMI value (BMIz) with the LMS method, the following formula [8] is used

$$BMIz = \left[\left(\frac{\mathrm{BMI}}{M}\right)^{\mathrm{L}} - 1\right] \div (\mathrm{L} \times \mathrm{S}) \quad \text{(Equation 1)}$$

Values for the L, M, and S are available on the CDC website for each sex and month of age. Values for the exponent, 'L' in equation #1 are negative at all ages, and range from -3.4 to -1.0.

Limitations of LMS-based z-scores

The LMS power transformation required to achieve normality constrains the maximum z-score that is obtainable at a given sex and age, making the LMS-derived z-scores ill-suited for identifying extreme values. No matter what the value of BMI is, the maximum obtainable z-score for a given sex/age is (-1) / (L × S). This occurs because as BMI becomes very large relative to M, $(\mathrm{BMI} \div \mathrm{M})^{\mathbf{L}}$ **in equation #1 approaches** 0. (L is always negative.) Given a very large BMI, the maximum obtainable z-score depends only on the values of L and S, not on BMI.

Consider a 200 month-old (16 years, 8 months) girl with a BMI of 33 $\mathrm{kg/m^2}$. The LMS parameters for this sex/age are L= -2.18, M= 20.76, and S= 0.148. Substituting these values into equation #1 gives,

$$BMIz = \left[\left(\frac{33.0}{20.76}\right)^{-2.18} - 1\right] \div (-2.18 \times 0.148)$$

= -0.636 / -0.323

= 1.97 (or the 97.6^{th} percentile)

Now, suppose that this BMI was incorrectly entered as 333 $\mathrm{kg/m^2}$. Her BMIz would now be calculated to be 3.1, which is close to the theoretical maximum z-score for this sex/age. This 300 $\mathrm{kg/m^2}$ difference (33 vs. 333) in the 2 BMIs correspond to only a 1.1 SD difference (1.97 vs. 3.08) between the 2 BMIz values. In addition to limiting the maximum z-score value that can be obtained at any sex/age, the LMS transformation results in the mapping of a wide range of very high BMIs into a narrow range of high z-scores.

This precludes the use of the LMS z-scores in the identification of values that are may be data errors. Modified z-scores were introduced to address this limitation.

Calculation of modified z-scores in the CDC growth charts

The SAS program for the CDC growth charts calculates these modified z-scores, but the following text explains the details of these calculation. In these modified z-scores, the BMI of a child is expressed relative to the median BMI in units of ½ of the distance between 0 and +2 z-scores. The following example shows the calculation of the modified z-score for the 200-month-old girl with a BMI of 333 kg/m².

First, one would calculate the BMI values associated with z-scores of 0 and 2 to obtain the distance that will be used in the calculation of the modified z-score. The formula to calculate the BMI corresponding to any z-score is

$$\mathrm{BMI} = \mathrm{M} \times [1 + \mathrm{LSz}]^{(1/L)} \text{ (Equation 2)}$$

in which 'z' represents the z-score of interest. For the modified z-score calculations, one would be interested in the BMIs associated with z-scores of 0 and 2. A z-score of 0 is simply the median at that sex/age, and in this case, the BMI is 20.76 kg/m². (When z=0, L × S × z is 0, and the BMI is equal to M in equation #2).

The BMI calculation for a z-score of 2 uses the L, M, and S values. For this 200-month-old girl, L = -2.18, M = 20.76, and S = 0.148. The BMI associated with a z-score of +2.0 would then be (from equation #2):

$$20.76 \times [1 + (-2.18 \times 0.148 \times 2.0)]^{-0.46} = 33.44 \text{ kg/m}^2.$$

One half of the distance between a z-score of 0 and +2 is (33.44 – 20.76) / 2 = 6.34 kg/m². This is the distance used in the calculation of the modified z-score.

The distance of the observed BMI (333 kg/m²) from the median BMI is expressed relative to this 6.34 kg/m² distance to yield the value of the modified z-score:

$$(333 - 20.76) \div 6.34 = \text{a modified z-score of } 49.2$$

(The value of 333 is 49.2 times the 6.34 kg/m² estimated distance from the median BMI.)

The use of modified z-scores identifies this value as very extreme (modified z-score of 49.2), whereas the use of the LMS z-score did not (z-score of 3.1)

Additional information

This modified z-score approach in the SAS program is somewhat similar to those used in the development of the 1977 NCHS/WHO growth charts [9] and in the WHO reference standards [10]. The extrapolation in the WHO standards, however, was based on the distance between 2 SD and 3 SD rather than on ½ of the distance between 0 and 2 z-scores (CDC method).

Based on these modified z-scores, extreme values of weight, height and BMI are flagged as BIVs and the cut points for these flags are explained in the SAS program for the CDC growth charts. These flagged values are likely to represent data errors, and if possible, should be further examined.

It should be realized, however, that very extreme values are not necessarily data errors. For example, there were six 15- to 19-year-olds examined in NHANES from 1999-2000 through 2013-2014 who had BMIs between 60 and 69 kg/m^2. The LMS-based z-scores of these six subjects ranged from 2.6 to 3.4, whereas their modified z-scores ranged from 6.2 to 8.7.

References

1. Freedman DS, Lawman HG, Skinner AC, McGuire LC, Allison DB, Ogden CL. Validity of the WHO cutoffs for biologically implausible values of weight, height, and BMI in children and adolescents in NHANES from 1999 through 2012. Am J Clin Nutr. 2015;102:1000–6.
2. Freedman DS, Lawman HG, Pan L, Skinner AC, Allison DB, McGuire LC, Blanck HM. The prevalence and validity of high, biologically implausible values of weight, height, and BMI among 8.8 million children. Obesity (Silver Spring). 2016;Mar 17.
3. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, Wei R, Curtin LR, Roche AF, Johnson CL. 2000 CDC Growth Charts for the United States: methods and development. Vital Health Stat 11. 2002;11:1–190.
4. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med. 1992;11:1305–19.
5. Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr. 1990/01/01. 1990;44:45–60.
6. Flegal KM, Wei R, Ogden CL, Freedman DS, Johnson CL, Curtin LR. Characterizing extreme values of body mass index-for-age by using the 2000 Centers for Disease Control and Prevention growth charts. Am J Clin Nutr. 2009;90:1314–20.
7. Flegal KM, Cole TJ. Construction of LMS parameters for the Centers for Disease Control and Prevention 2000 growth charts. Natl Health Stat Report. 2013;9:1–3.
8. Cole T, Bellizzi M, Flegal K, Dietz W. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ. 2000/05/08. 2000;320:1240–3.
9. Dibley MJ, Goldsby JB, Staehling NW, Trowbridge FL. Development of normalized curves for the international growth reference: historical and technical considerations. Am J Clin Nutr. 1987;46:736–48.
10. World Health Organization. Department of Nutrition for Health and Development. WHO Child Growth Standards. Length/height-for-age, weight-for-age, weight-for-length, weight-for-height, and body mass index-for-age. Methods and Development. Geneva; 2006. pages 301-304