Modified z-scores in the CDC growth charts

Data quality assessment on anthropometry data:

When analyzing 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). While these extreme values may represent errors in data entry or measurement, they may 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 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 ten smoothed percentile (between the 3rd and 97th) of BMI for 24 to 240 month-old children and adolescents [3]. These smoothed estimates were then 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, the approach used in the CDC growth charts to estimate these 3 parameters differs substantially from that proposed by Cole [6,7]. In the CDC method, the LMS parameters for BMI were derived from the equations for 10 smoothed percentiles, whereas the original method derived these parameters from the entire population.

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

used

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

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 obtainable at a given sex and age, making the LMS-derived z-scores ill-suited for identifying extreme values. No matter the BMI value, the maximum possible z-score for a given sex/age is $(-1) / (L \times S)$. This occurs because as BMI becomes very large relative to M, $(BMI \div M)^{L}$ in equation #1 approaches 0 as L is less than -1 at all ages. Given a very large BMI, the maximum obtainable z-score depends largely 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 kg/m². 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,

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

$$= -0.636 / -0.323$$

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

Now, suppose that this BMI was incorrectly entered as 333 kg/m². Her BMIz would then be calculated as 3.1, a value close to the theoretical maximum z-score for this sex/age. This 300 kg/m² difference (33 vs. 333) between the 2 BMIs corresponds 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 obtainable at any sex/age, the LMS transformation results in mapping a wide range of very high BMIs into a

narrow range of high z-scores.

This precludes the use of the LMS z-scores to identify values that may be data errors. Therefore, 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, and the following text explains the details of these calculations. 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 the median BMI and the BMI that has a z-score value of +2 or -2 SDs. The first 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 used to calculate the modified z-score. The formula to calculate the BMI corresponding to any z-score is

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

in which 'z' represents the z-score of interest. 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.

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):

$$20.76 \times [1 + (-2.18 \times 0.148 \times 2.0)]^{-0.46} = 33.40 \text{ kg/m}^2.$$

One-half of the distance between a z-scores of 0 and +2 is (33.40 − 20.76) / 2 = 6.32 kg/m² ; this is the SD distance used in the calculation of the modified z-score.

The distance of the observed BMI (333 kg/m^2) from the median BMI is expressed relative to this 6.34 kg/m^2 distance to yield the value of the modified z-score:

$$(333 - 20.76) \div 6.32 = 49.2 \text{ SDs}$$

The use of modified z-scores identifies this value as very extreme (modified z-score of 49.2), whereas its LMS z-score was not ($z = 3.1$)

The second example shows the calculation of the modified z-score for a 200-month-old girl with a BMI below the median. The first step is to calculated the BMI value at −2 SDs using the L, M, and S parameters:

$$20.76 \times [1 + (-2.18 \times 0.148 \times -2.0)]^{-0.46} = 16.52 \text{ kg/m}^2.$$

If this girl had a BMI of 12, the modified BMIz value would then be

$$(12.0 - 20.76) \div (0.5 * (20.76 - 16.52)) = -4.1$$

In a program, both positive and negative values of modified BMIz can be calculated with the following steps:

    If BMI > M, SD_distance = 0.5 * ((M * (1 + L*S*2) ^ (1/L)) − M)
    If BMI < M, SD_distance = 0.5 * (M − ((M * (1 + L*S * (−2)) ^ (1/L))))
    The modified BMIz would then be calculated as (BMI − M) ÷ SD_distance

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 may represent data errors and should be further examined if possible.

However, it should be realized that extreme values are not necessarily data errors. For example, 15 2- to 19-year-olds examined in NHANES from 1999-2000 through 2017-2018 had a modified BMIz > 8. The highest modified BMIz value was 11.4 SDs, corresponding to a 3-year-old boy with a BMI of 31.6 kg/m^2.

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. The American Journal of Clinical Nutrition 2015;102:1000–6.

[2] Freedman DS, Lawman HG, Pan L, Skinner AC, Allison DB, McGuire LC, et al. The prevalence and validity of high, biologically implausible values of weight, height, and BMI among 8.8 million children. Obesity (Silver Spring, Md) 2016;Mar 17. https://doi.org/10.1002/oby.21446.

[3] Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. 2000 CDC Growth Charts for the United States: methods and development. Vital and Health Statistics Series 11, Data from the National Health Survey 2002;11:1–190.

[4] Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Statistics in Medicine 1992;11:1305–19.

[5] Cole TJ. The LMS method for constructing normalized growth standards. European Journal of Clinical Nutrition 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. American Journal of Clinical Nutrition 2009;90:1314–20. https://doi.org/10.3945/ajcn.2009.28335.

[7] Flegal KM, Cole TJ. Construction of LMS parameters for the Centers for Disease Control and Prevention 2000 growth charts. National Health Statistics Reports 2013;9:1–3. https://doi.org/10.1371/journal.pone.0101791.

[8] Cole T, Bellizzi M, Flegal K, Dietz W. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ (Clinical Research Ed) 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. The American Journal of Clinical Nutrition 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.