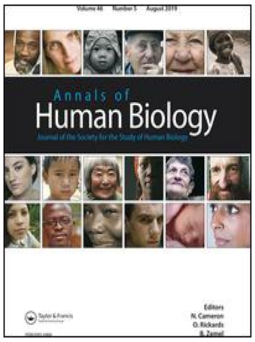

Annals of Human Biology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iahb20

A method for calculating BMI z-scores and percentiles above the 95th percentile of the CDC growth charts

Rong Wei , Cynthia L. Ogden , Van L. Parsons , David S. Freedman & Craig M. Hales

ANNALS OF HUMAN BIOLOGY
https://doi.org/10.1080/03014460.2020.1808065

RESEARCH PAPER

A method for calculating BMI z-scores and percentiles above the 95th percentile of the CDC growth charts

Rong Wei[a], Cynthia L. Ogden[a], Van L. Parsons[a], David S. Freedman[b] and Craig M. Hales[a]

[a]National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, USA; [b]Division of Nutrition, Physical Activity, and Obesity, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA

ABSTRACT

Background: The 2000 CDC growth charts are based on national data collected between 1963 and 1994 and include a set of selected percentiles between the 3rd and 97th and LMS parameters that can be used to obtain other percentiles and associated z-scores. Obesity is defined as a sex- and age-specific body mass index (BMI) at or above the 95th percentile. Extrapolating beyond the 97th percentile is not recommended and leads to compressed z-score values.

Aim: This study attempts to overcome this limitation by constructing a new method for calculating BMI distributions above the 95th percentile using an extended reference population.

Subjects and methods: Data from youth at or above the 95th percentile of BMI-for-age in national surveys between 1963 and 2016 were modelled as half-normal distributions. Scale parameters for these distributions were estimated at each sex-specific 6-month age-interval, from 24 to 239 months, and then smoothed as a function of age using regression procedures.

Results: The modelled distributions above the 95th percentile can be used to calculate percentiles and non-compressed z-scores for extreme BMI values among youth.

Conclusion: This method can be used, in conjunction with the current CDC BMI-for-age growth charts, to track extreme values of BMI among youth.

ARTICLE HISTORY
Received 24 April 2020
Revised 6 July 2020
Accepted 13 July 2020

KEYWORDS
Childhood obesity; body mass index; growth percentiles; Half Normal distribution

Introduction

Growth charts are widely used in research and clinical practice to assess childhood growth and development. Body mass index (BMI)-for-age growth charts are used to define and monitor obesity among children and adolescents. In the United States, obesity among children and adolescents 2–19 years is defined as at or above the sex- and age-specific 95th percentile of the 2000 CDC BMI-for-age growth charts (Kuczmarski et al. 2002; Ogden and Flegal 2010).

The 2000 CDC BMI-for-age growth charts include selected smoothed percentiles between the 3rd and 97th percentiles. The sample size used to develop the charts was not adequate for the calculation of percentiles outside of this range. Parameters based on the skewness (L), median (M), and coefficient of variation (S) from the smoothed percentiles of BMI were calculated to match the selected smoothed percentiles between the 3rd and 97th percentile so that percentiles and z-scores can be calculated for any BMI in this range (Cole 1990; Cole and Green 1992; Flegal and Cole 2013).

Although percentiles and z-scores beyond the 97th percentile can be extrapolated using the LMS parameters $Z = [(\mathrm{BMI}/M)L - 1]/(LxS)$; $L \neq 0$, their use is discouraged because the extrapolated values do not reflect the sparse underlying data (Flegal et al. 2009; Flegal and Cole 2013). Moreover, the z-scores obtained through extrapolation are compressed so that large differences in BMI are reflected as small differences in z-scores that distort clinically meaningful changes (Woo 2009; Freedman, Butte, Taveras, Goodman, Ogden, et al. 2017; Freedman, Butte, Taveras, Lundeen, et al. 2017). For example, two 10-year-old boys with extremely high BMIs that are substantially different, 50 and 55, would have BMI z-scores that differ by only 0.03 units (2.87 vs. 2.90). This is because these BMI z-score values are close to the maximum z-score that is possible for 10-year-old boys.

Despite the caution against using the CDC growth chart LMS parameters to calculate percentiles and z-scores beyond the 97th percentile, expressing extreme BMI values using LMS-derived z-scores is common practice in intervention studies (US Preventive Services Task Force (USPSTF) et al. 2017). Given this limitation, another method is needed to assess changes in extreme values of BMI in youth. This study describes a method developed using nationally representative data from 1963–2016 for children and adolescents aged 2–19 years at or above the 95th percentile of BMI-for-age to produce an extended BMI reference population beyond the original BMI reference population. Data for children and

CONTACT Rong Wei rrw5@cdc.gov National Center for Health Statistics, Centers for Disease Control and Prevention, 3311 Toledo Rd, Hyattsville, MD 20782, USA

Supplemental data for this article is available online at https://doi.org/10.1080/03014460.2020.1808065.

adolescents at or above the 95th percentile, instead of the 97th percentile, were used to ensure an adequate sample for estimation of the model. The new BMI metric, which alters only BMI percentiles and z-scores that are above the 95th percentile, can be combined with the current BMI percentiles and z-scores to form a single metric that can be used across the entire BMI distribution.

Methods

The National Health and Nutrition Examination Survey (NHANES) is a serial cross-sectional survey of the non-institutionalised US population. The survey includes an in-home interview followed by a physical examination and laboratory testing performed at mobile examination centres. The physical exam included standardised measurements of height and weight (CDC National Centre for Health Statistics 2020).

Data from the following NHANES surveys were used to create the 2000 CDC BMI-for-age growth charts for children and adolescents aged 2 to 19 years (24 to 239 months): the precursor to NHANES the National Health Examination Survey (NHES) II (1963–1965, ages 6–12.5 years), NHES III (1966–1970, ages 12–18.5 years), NHANES I (1971–1974, ages 2–19 years), NHANES II (1976–1980, ages 2–19 years), and NHANES III (1988–1994, ages 2–5 years) (Kuczmarski et al. 2002). For this study, respondents from NHANES III (ages 6–19 years) and continuous NHANES 1999–2000 to 2015–2016 (ages 2–19 years) were included in this analysis to expand the previous growth chart sample (CDC National Centre for Health Statistics 2020). Those with a BMI at or above the sex- and age-specific 95th percentile of the 2000 CDC growth charts (Centres for Disease Control and Prevention (CDC 2020a)) were included in the extended reference population. An initial effort to include data from children and adolescents at or above the 97th percentile was not successful as the small sample size was too small. Pregnant girls in NHANES III and continuous NHANES were excluded.

With the extended method, BMI z-scores and percentiles up to the 95th percentile are calculated in the usual way using the LMS parameters, but z-scores and percentiles above the 95th percentile are calculated using an extended reference population that combines children and adolescents with obesity (BMI $\geq$95th percentile) from the 2000 growth chart reference population together with those from more recent NHANES surveys (see above). The extended method proposes a recalculation of all percentiles above the 95th percentile because it allows the use of a larger reference population, or sample size, for estimating BMI distributions at very high BMI levels than does using the 97th percentile (final analytic sample size = 8777 versus 6082).

The ability to estimate all percentiles using one distribution up to the 95th percentile and a separate distribution above the 95th percentile is given by the property of conditional probability:

$$P\left(X \geq \mathrm{BMI}_{p^{\mathrm{th}}}\right) = P\left(X \geq \mathrm{BMI}_{p^{\mathrm{th}}} | X \geq \mathrm{BMI}_{95^{\mathrm{th}}}\right) \cdot P\left(X \geq \mathrm{BMI}_{95^{\mathrm{th}}}\right), \tag{1}$$

for $p > 95$, where X is a random variable for BMI values in a population, $BMI_{p^{th}}$ is the BMI at a given sex- and age-specific p^{th} percentile. $P\left(X \geq BMI_{p^{th}}\right)$ is the probability that BMI values in a population are at or above a given $BMI_{p^{th}}$. For example, an 8-year-old girl with a BMI of 17 is at the 75th percentile (according to the LMS parameters); therefore, $P(X \geq 17) = 1 - 0.75 = 0.25$. An 8-year-old girl with a BMI of 26 is above the 95th percentile (the BMI value at the 95th percentile for a girl 96.5 months of age is 20.7), so Equation (1) would apply. The extended method uses a combination of 2 underlying distributions to calculate $P(X \geq 26)$ and hence the percentile and z-score. First, $P\left(X \geq BMI_{95^{th}}\right)$ is the probability that BMI is at or above the 95th percentile using the LMS parameters and equals 0.05. $P\left(X \geq BMI_{p^{th}} | X \geq BMI_{95^{th}}\right)$ is the conditional probability that BMI in a population is at or above a certain BMI at the p^{th} percentile given that the BMI is at or above the 95th percentile, which is the case for the BMI of 26. This conditional probability can be estimated using a separate distribution of BMI in a reference population of children and adolescents who have a BMI at or above the 95th percentile. Once $P\left(X \geq 26 | X \geq BMI_{95^{th}}\right)$ is estimated, it is multiplied by 0.05 to find $P(X \geq 26)$ for the full population and hence the percentile for an 8-year-old girl with a BMI of 26.

The remaining task is to find a probability distribution that represents the distribution of BMI above the 95th percentile and estimate the associated parameters. To do this, we divided the extended sample of children and adolescents with BMI at or above the 95th percentile from 1963–2016 into 36 6-month age intervals for boys and girls separately (72 total), as was done in the development of the 2000 CDC growth charts (Kuczmarski et al. 2002). The half-normal distribution was considered as a reasonable candidate for fitting as it is based on features of a normal distribution; in particular, the shape of a standard half-normal distribution corresponds to the shape of the standard normal restricted to the positive axis (Wei et al. 2019). A successful fit leads to simple computational forms for percentiles and z-scores. This scenario describes a reasonable underlying model for the empirical distribution of BMI values at each age interval in the current study, where only values *above* the sex- and age-specific 95th percentile are considered.

The half-normal distribution, with lower bound 0, is characterised by a single parameter - the shape parameter σ - which is proportional to both its mean, $\sigma\sqrt{\frac{2}{\pi}}$, and standard deviation, $\sigma\sqrt{\left(1-\frac{2}{\pi}\right)}$, as well as the standard deviation of the corresponding 0-centred normal distribution and for the purpose at hand, describes the "spread" of the distribution. For BMI values above the 95th percentile, the dispersion generally increases with age and is different between boys and girls; therefore, the shape parameter σ was estimated for each 6-month age interval for boys and girls separately.

At each interval a, $\hat{\sigma}_a$ was estimated by the form:

$$\hat{\sigma}_a = \sqrt{\frac{\pi}{2}} \frac{\sum_{i=1}^{n_a} w_i\left(\mathrm{BMI}_i - \mathrm{BMI}_{95\mathrm{th}a}\right)}{\sum_{i=1}^{n_a} w_i}.$$

In this equation, BMI_i is the individual BMI, n_a is the unweighted sample size in interval a, w_i is the individual sample weight, and BMI_{95tha} is the BMI at the sex- and age-specific 95th percentile in interval a (See the Appendix for additional details). An approximate global design effect (Wei et al. 2019) on interval a, or variance inflation factor for an estimate due to the survey sample design, was estimated as

$$deff_a = 1 + [cv(\text{weights}_a)]^2,$$

where $cv(\text{weights}_a)$ is the coefficient of variation of the sample weights in interval a. The effective sample size was calculated as the actual sample size divided by the design effect, i.e.

$$n_{eff\,a} = \frac{n_a}{deff_a}.$$

The fit of the half-normal cumulative distribution function (CDF) was evaluated by plotting the empirical CDF of $(\text{BMI}-\text{BMI}_{95\text{th}a})$ along with its 95% confidence intervals at each 6-month interval and then overlaying the estimated CDF of a half-normal distribution based on parameter $\sigma_a = \hat{\sigma}_a$.

To reduce variation in $\hat{\sigma}_a$ across 6-month intervals and to facilitate the use of the modelled $(\text{BMI}-\text{BMI}_{95\text{th}})$ distributions for analysis, the estimates $\hat{\sigma}_a$ were smoothed by a weighted least squares polynomial regression for boys and girls separately using the effective sample size $n_{eff\,a}$ as the weight. The smoothed $\hat{\sigma}_a$ were estimated using the quadratic function

$$\hat{\sigma}_{a,\ \text{smooth}} = \hat{c}_0 + \hat{c}_1\text{age} + \hat{c}_2\text{age}^2 \tag{2}$$

with age expressed in years at the midpoints of each 6-month interval, i.e. 2.25, 2.75, 3.25, … . Other smoothing techniques, such as robust locally weighted regression (Cleveland 1979), gave very similar estimates for $\hat{\sigma}_{a,\ \text{smooth}}$.

Now, for any age, the distribution of BMI above the 95th percentile can be specified as a half-normal distribution with shape parameter $\hat{\sigma}_{a,\ smooth}$. The conditional expression from Equation (1), $P(X \geq \text{BMI}_{p^{th}} | X \geq \text{BMI}_{95^{th}})$, can be calculated using the CDF of the half-normal distribution at any 6-month interval as

$$P(X \geq \text{BMI}_{p^{th}} | X \geq \text{BMI}_{95^{th}}) = 2\left[1 - \Phi\left(\frac{\text{BMI}-\text{BMI}_{95\text{th}a}}{\hat{\sigma}_{a,\ \text{smooth}}}\right)\right],$$

where Φ is the CDF of the standard normal distribution. Now, the complete Equation (1) can be calculated as:

$$\begin{aligned} P(X \geq \text{BMI}_{p^{th}}) &= P(X \geq \text{BMI}_{p^{th}} | X \geq \text{BMI}_{95^{th}}) \cdot P(X \geq \text{BMI}_{95^{th}}) \\ &= 2\left[1 - \Phi\left(\frac{\text{BMI}-\text{BMI}_{95\text{th}a}}{\hat{\sigma}_{a,\ \text{smooth}}}\right)\right](0.05) \\ &= \left[1 - \Phi\left(\frac{\text{BMI}-\text{BMI}_{95\text{th}a}}{\hat{\sigma}_{a,\ \text{smooth}}}\right)\right](0.1). \end{aligned} \tag{3}$$

The BMI percentiles, for those larger than the 95% percentile, are typically expressed as 1 minus Equation (3). In percentage form Equation (3) can be rewritten as:

$$\text{BMI\%ile} = 90 + 10\Phi\left(\frac{\text{BMI}-\text{BMI}_{95\text{th}a}}{\hat{\sigma}_{a,\ \text{smooth}}}\right), \tag{4}$$

where BMI%ile is the BMI percentile. Any BMI percentile $\geq$95th percentile can be calculated using Equation (4); there is no upper limit. As a simple example, if $\text{BMI} = \text{BMI}_{95\text{th}}$, then $10\Phi(0) = 5$, and $\text{BMI\%ile} = 95$. Conversely, the BMI value for any percentile above the 95th percentile can be calculated by rearranging Equation (4):

$$\text{BMI} = \Phi^{-1}\left(\frac{\text{BMI\%ile}-90}{10}\right)\hat{\sigma}_{a,\ \text{smooth}} + \text{BMI}_{95\text{th}}, \tag{5}$$

where Φ^{-1} is the inverse of the standard normal distribution CDF. Any percentile derived using Equation (4) can be converted to a z-score (BMI_z) as:

$$\text{BMI}_z = \Phi^{-1}\left(\frac{\text{BMI\%ile}}{100}\right). \tag{6}$$

The supplemental material contains an R function and SAS programme that calculate these z-scores and percentiles based on the half-normal distribution and $\hat{\sigma}_{a,\ \text{smooth}}$ for children with a BMI at or above the 95th percentile of the CDC growth charts. These values among children with obesity are then combined with the current BMI z-scores and percentiles from the CDC growth charts (Centres for Disease Control and Prevention (CDC 2020b)) for children with lower BMI levels to form metrics that we refer to as "extended BMIz" and "extended BMI percentile."

Results

There were 8,831 children and adolescents aged 2–19 years from all surveys (1814 children from the original growth chart sample and 7017 additional children from 1988–2016) with BMI at or above the 95th percentile. Of these, 54 (0.61%) were excluded due to pregnancy, leaving a sample size of 8,777.

Table 1 shows the unweighted sample size by survey; 1,814 children and adolescents were the same as those in the reference population for the 2000 CDC BMI-for-age growth charts, and 6,963 were added from subsequent surveys. Table 2 shows unweighted sample sizes (mean = 120, range = 55 to 188), and effective sample sizes (mean = 2.12,

Table 1. Unweighted sample size for children and adolescents aged 2–20 years at or above the 95th percentile of 2000 CDC BMI-for-age growth chart, by survey.

Survey	Ages (year)*	n
Original CDC growth charts		
NHES II 1963–65	6 – 12.5	320
NHES III 1966–70	12 – 18.5	322
NHANES I 1971–74	2 – 20	379
NHANES II 1976–78	2 – 20	396
NHANES III 1988–94	2 – 6	397
Supplemental data		
NHANES III 1988–94	6 – 20	863
NHANES 1999–2000	2 – 20	738
NHANES 2001–02	2 – 20	726
NHANES 2003–04	2 – 20	722
NHANES 2005–06	2 – 20	802
NHANES 2007–08	2 – 20	603
NHANES 2009–10	2 – 20	635
NHANES 2011–12	2 – 20	593
NHANES 2023–14	2 – 20	629
NHANES 2015–16	2 – 20	652
Total		8777

*Exact age range is: number ≤ age < number, e.g. 2–20 means 2 ≤ age < 20.
NHES: National Health Examination Survey.
NHANES: National Health and Nutrition Examination Survey.

Table 2. Unweighted and effective sample size for children and adolescents aged 2–20 years at or above the 95th percentile of 2000 CDC BMI-for-age growth charts, by sex and age, 1963–2016.

	Males			Females		
Age (months)	*N* Growth charts	*N* Suppl.* data	Effective sample size	*N* Growth charts	*N* Suppl.* data	Effective sample size
2–2.49	25	30	29.2	28	49	42.0
2.5–2.99	36	60	49.2	33	58	43.0
3.0–3.49	41	53	36.6	45	40	32.9
3.5–3.99	43	53	41.9	39	49	43.8
4.0–4.49	43	70	55.4	43	59	45.6
4.5–4.99	43	85	68.8	43	59	47.9
5.0–5.49	41	70	49.0	48	60	43.7
5.5–5.99	45	56	44.1	56	62	54.1
6.0–6.49	20	86	52.4	15	82	50.6
6.5–6.99	18	91	55.9	15	73	45.6
7.0–7.49	22	98	63.1	20	91	53.0
7.5–7.99	16	108	64.8	16	100	72.0
8.0–8.49	18	116	52.6	20	108	68.3
8.5–8.99	14	121	65.2	25	102	54.0
9.0–9.49	26	108	76.2	26	107	66.3
9.5–9.99	28	122	78.6	32	106	68.0
10.0–10.49	22	122	73.9	25	81	58.4
10.5–10.99	17	110	63.0	24	101	58.3
11.0–11.49	23	122	69.9	26	117	74.4
11.5–11.99	35	133	84.8	24	133	83.3
12.0–12.49	22	130	72.9	26	119	62.8
12.5–12.99	25	121	72.1	32	156	80.3
13.0–13.49	24	137	71.5	32	119	68.1
13.5–13.99	31	100	62.3	25	137	68.1
14.0–14.49	23	118	58.8	23	120	69.3
14.5–14.99	25	116	66.5	22	109	54.5
15.0–15.49	26	121	68.7	26	106	53.4
15.5–15.99	20	92	54.8	21	98	46.6
16.0–16.49	19	124	69.4	25	108	63.2
16.5–16.99	13	137	76.5	24	90	54.6
17.0–17.49	20	108	62.3	25	114	49.2
17.5–17.99	17	101	55.8	15	81	39.2
18.0–18.49	15	106	58.5	11	106	55.4
18.5–18.99	7	100	48.0	5	93	43.5
19.0–19.49	11	92	54.2	6	90	43.3
19.5–19.99	9	79	43.1	10	84	48.3

*Suppl., Supplemental. The examination year and ages of children included in the original CDC growth charts and in the supplemental data are shown in Table 1.

range 1.61 to 2.82), and effective sample sizes (mean = 57.98, range = 21.98 to 84.77) by sex and 6-month age interval. Figure 1 shows the BMI values of all children and adolescents in the sample by sex and age and illustrates the increasing dispersion of the BMI distribution with increasing age.

Plots of the empirical CDF of $(\mathrm{BMI}-\mathrm{BMI}_{95\mathrm{th}a})$ overlaid with the fitted half-normal CDF at each 6-month interval (Supplemental figures) show that the half-normal distribution estimated with $\hat{\sigma}_a$ fits the data well for most intervals, although some lack of fit is expected among the 72 intervals given the small effective sample sizes for some sex-age groups.

Figure 2 shows plots of $\hat{\sigma}_a$ for each 6-month interval and $\hat{\sigma}_{a,\mathrm{smooth}}$ calculated using weighted polynomial regression for boys and girls separately. R^2 of 89.3% for boys and 88.2% for girls indicate a good fit of the regression models. The regression parameter estimates are shown in Table 3.

Figure 3 shows extended and original LMS-generated BMI z-scores 2 through 4 (z-2, z-3, z-4) overlaid on the observed data used to generate extended z-scores. The original z-4 curves have a vertical asymptote around age 6 in boys and age 5 in girls, and the original z-3 curve behaves erratically, while the extended z-2 through z-4 curves increase in a manner that is consistent with growth indicated by the percentile curves below the 95th percentile. The z-score of 1 is not included in the figure as it falls below the 95th percentile. As an example calculation, for a 12-year-old (144 months) boy with a BMI of 40, $\hat{\sigma}_{a,smooth}$ is calculated as 5.3 using the regression coefficients in Table 3 and formulas (2) above: $0.3728 + 0.5196{*}\mathrm{age} - 0.0091{*}\mathrm{age}^{\mathbf{2}}$. Based on Equation (4), the estimated percentile is 99.985 $(90 + 10{*}\Phi((40 - 24.23)/5.3))$ and the estimated z-score, from Equation (6), is 3.6 $(\mathrm{BMIz} = \Phi^{-\mathbf{1}}(99.985/100))$. Using the original L, M and S parameters, the z-score for this boy is 2.7.

To compare the extended method with the original growth charts, Figure 4 shows the 97th percentile from the original charts along with the 97th percentile generated using the extended method. The original and extended 97th percentiles are very similar, although the 97th percentile from the extended method is slightly higher than the original 97th percentile. The higher values of the extended percentiles are not unexpected given the addition of newer data that reflect the increase in obesity prevalence among youth in the US over the last 30 years (Fryar et al. 2018). The increase has been slightly larger in boys, which explains why

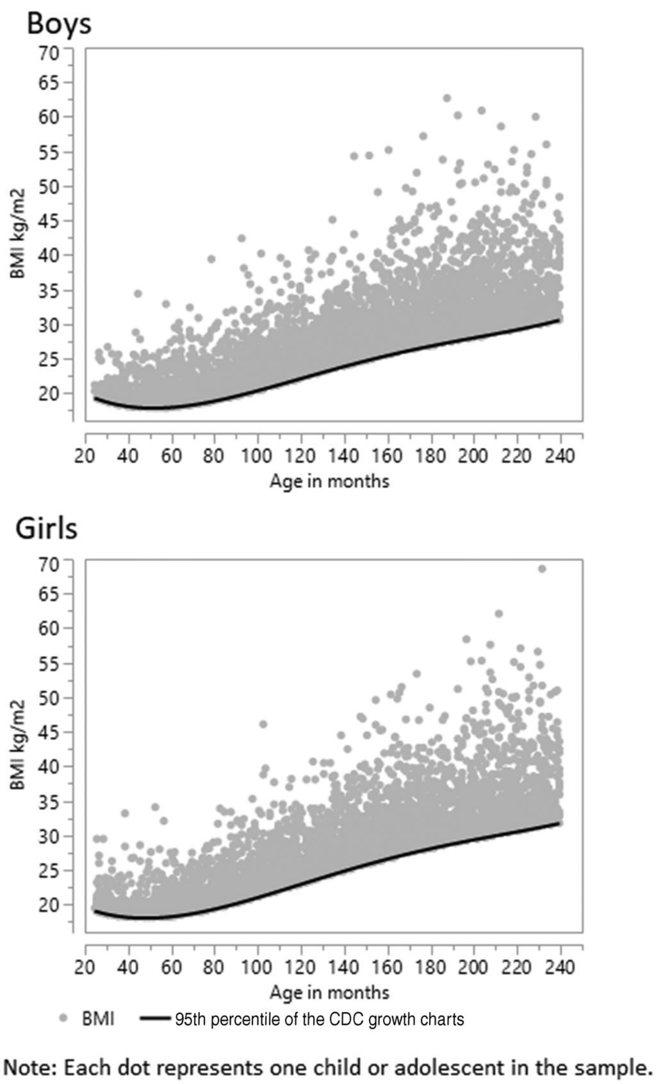

Figure 1. Age and body mass index (BMI) distribution of children and adolescents aged 2–20 years at or above the 95th percentile of 2000 CDC BMI-for-age growth chart, by sex, 1963–2016.

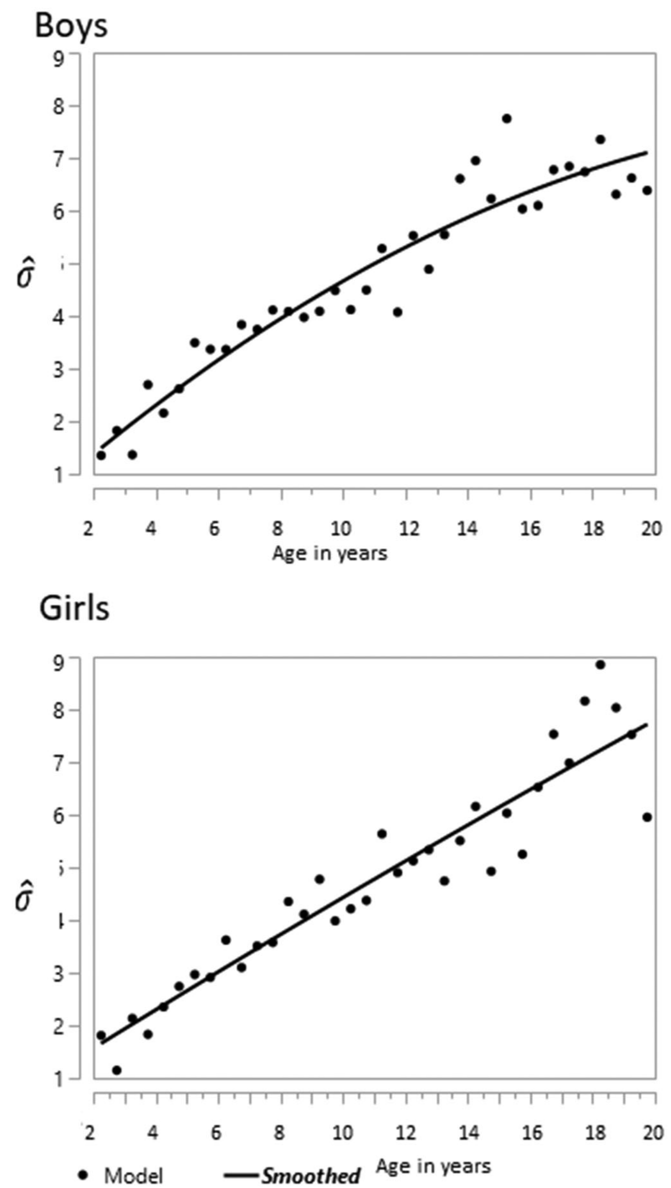

Figure 2. Sex-specific estimates of dispersion parameter σ characterising the half-normal distribution of BMI above the sex- and age-specific BMI 95th percentile for each 6-month interval from age 2 to 20 years and the smoothed.$\hat{\sigma}$

there is a larger difference between the extended 97th percentile and the CDC 97th percentile in boys, particularly in older boys than in girls. Moreover, this will result in calculated extended z-scores being slightly lower than those based on the original growth charts for BMI values just above the 95th percentile. For example, a 16-year-old boy with a BMI of 29.3 would be at the 97th percentile and have a BMIz of 1.88 on the original scale. Because the extended 97th percentile is about 1.6 kg/m^2 higher (30.9), this 16-year-old boy would have an extended BMIz of 1.76, a value that is 0.12 SDs lower than on the original BMIz scale.

Table 3. Regression parameters for dispersion parameter $\hat{\sigma}_{smooth}$* characterising the half-normal distribution of BMI above the sex- and age-specific BMI 95th percentile for each 6-month interval from age 2 to 20 years.

	Parameter			
Sex	Intercept ($\hat{c}_0$)	Age ($\hat{c}_1$)	Age^2 ($\hat{c}_2$)	R^2
Boys	0.3728	0.5196	−0.0091	89.31%
Girls	0.8334	0.3712	−0.0011	88.15%

*Estimated using $\hat{\sigma}_{smooth} = \hat{c}_0 + \hat{c}_1 age + \hat{c}_2 age^2$, with age expressed in years.

Discussion and summary

The extended method was developed using half normal distributions to model the distribution of BMI above the 95th percentile. This alleviates the problems of very high BMIs being compressed into a narrow range of z-scores with maximum values when extrapolating the CDC-derived LMS parameters for BMIs above the 97th percentile (Kuczmarski et al. 2002; Flegal et al. 2009; Flegal and Cole 2013). The problem arose because the LMS parameters (Cole 1990; Cole and Green 1992) were estimated from the skewed BMI distribution and from the fact that only percentiles between the 3rd and the 97th were used because of sparse data at the extreme (Kuczmarski et al. 2002). This skewness is indicated by the L parameter (power transformation for normality) being far smaller than 1 (where 1 indicates no transformation) and between −2 and

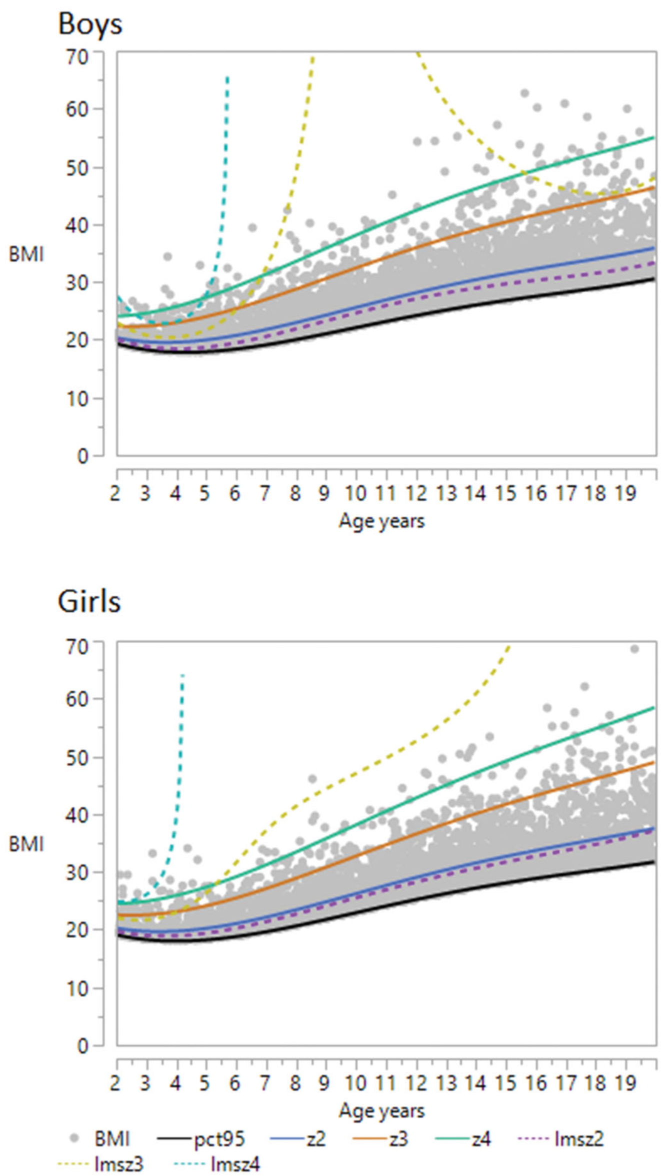

Figure 3. Extended and original LMS generated BMI z-scores 2 through 4 overlaid on the additional data points used to generate extended z-scores.
Z2, z3, and z4 refer to extended BMIz, while lmsz2, lmsz3, and lmsz4 refer to the current CDC BMI z-scores based on the LMS parameters.

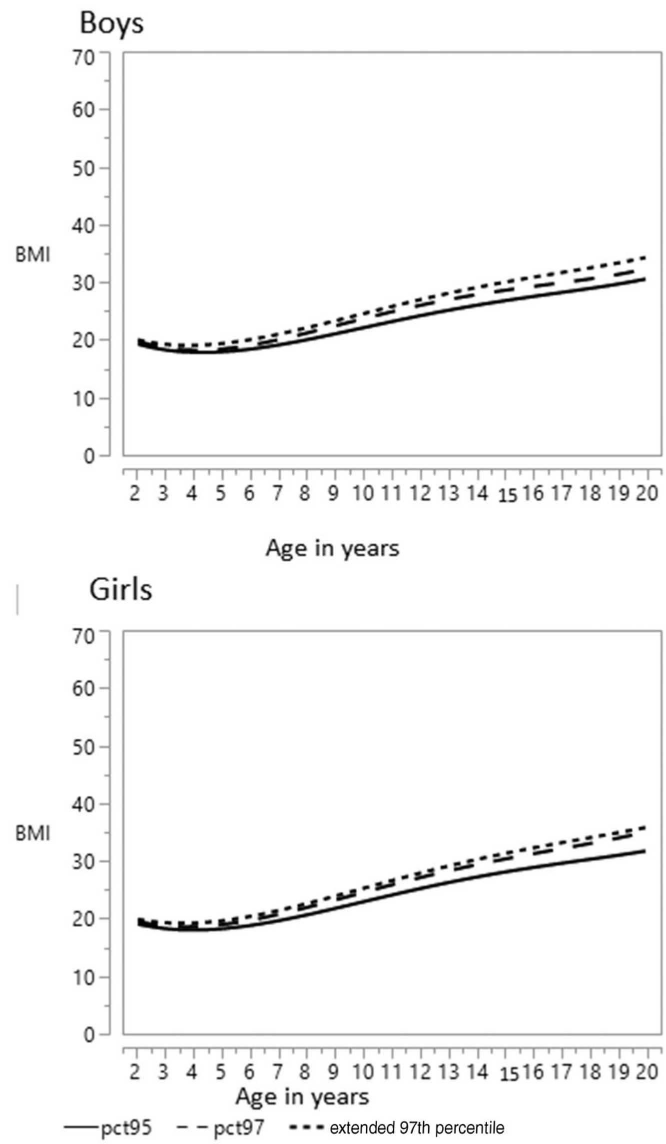

Figure 4. Comparison of original and extended 97th percentile.

−3 at most ages in the CDC growth charts (Centres for Disease Control and Prevention (CDC 2020a)).

These low values of the L parameter lead to the upper tail of the BMI distribution being compressed into a narrow z-score range at most ages (Freedman, Butte, Taveras, Goodman, Ogden, et al. 2017; Freedman, Butte, Taveras, Lundeen, et al. 2017) with an upper limit for BMI z-score that varies substantially by age and sex (Woo 2009). This compression can result in similarly aged children with markedly different BMIs having similar z-scores. In contrast, the half-normal distribution is characterised by a single parameter - the shape parameter σ - which is analogous to the standard deviation of the normal distribution and describes the variance or "spread" of the distribution. The use of the half-normal distribution for BMIs $\geq$ 95th percentile allows BMIs for children with obesity to be mapped to a percentile and z-score and avoids the problems of very high BMIs compressed into a narrow range of z-scores. With this method, extreme BMI and changes in BMI can be conveyed using a meaningful and familiar metric.

It has been suggested that applying the LMS method to the data in the CDC growth charts rather than deriving the parameters from already smoothed percentiles in the CDC approach (Cole 2010) might alleviate the compression of very high BMIs into a narrow z-score range. However, modelling the BMI distribution using the LMS procedure on all the data in the CDC growth charts resulted in only minor changes in the values of L (skewness) parameter and in the calculated z-scores from the current values (Centres for Disease Control and Prevention (CDC 2020b)). Further, as assessed by worm plots and Q statistics, the LMS distribution fit the CDC growth charts data poorly. Although the LMS method has been widely used, the WHO growth standards used this method only for z-scores between −3 and +3; more extreme z-scores were calculated by extending the distance between 2 and 3 SDs (or between −2 and −3 SDs) outwards (World Health Organization. Department of Nutrition for Health and Development 2006).

The extended method has some advantages. First, the method is based, in part, on more recent population data and the patterns reflect BMI in current children and adolescents. Second, unlike other methods, such as adjusted percent distance from the median (Freedman et al. 2019 Aug 23), this method allows investigators to continue working with the familiar z-score scale. Another advantage is that the z-scores and percentiles can be used directly in conjunction with percentiles and z-scores below the 95th percentile, and expressing BMI levels on the extended scale would not alter the surveillance of secular trends in obesity. Third, although there are some minor discrepancies between the original CDC chart BMI-for-age curves and the described method between the 95th to 97th percentiles, the extended method can be used to track extreme BMI values without affecting the original charts below the 95th percentile. By definition, the 95th percentiles are identical, as are percentiles (and z-scores) below the 95th percentile (z-score < 1.645). Finally, the method could be incorporated into SAS, R, and other programmes so that the transition would be seamless to the user. An R function and SAS programme are included in the supplemental material.

The extended method has some limitations. First, data from the 1960s and 1970s were combined with more recent data through 2016 when the prevalence of obesity has increased. BMI values above the 95th percentile from children in the more recent period have shifted to the right. For example, the mean BMI of a 16-year-old girl with obesity in the original growth charts data was 33.6, while the comparable mean BMI in the newer data was about 1 kg/m^2 higher (34.7). Second, the half-normal assumption for BMI above the 95th percentile may not accurately reflect the distribution for every sex and 6-month age group, and this was evident (supplemental figures) in a few of the 72 groups, such as boys who were between 42 and 47 months of age (Wei et al. 2019). Third, the extended calculation of z-score introduces a difference in the reference population between the 95th and 97th percentiles where many of the extended z-scores (and percentiles) will be slightly lower than those based on the original growth charts; this may affect the assessment of change where BMI crosses the 95th percentile. For example, a 10.0-year-old boy with a BMI of 21.6 has a BMI percentile of 93.9. If this boy had a BMI of 23.6 at age 11.0 y, his BMI percentile based on the current growth charts would be 95.6. However, his extended BMI percentile would be slightly lower at 95.3. In both cases, he would be considered to have obesity.

The utility of the extended method to measure changes in extreme BMI over time will be better understood with its implementation in intervention studies. The compression in the upper tail of the original LMS-based z-score, along with the differences in the maximum attainable z-score across sex and age, makes the original z-scores unsuitable for this purpose (Woo 2009; Freedman, Butte, Taveras, Goodman, Ogden, et al. 2017; Freedman, Butte, Taveras, Goodman, and Blanck 2017; Freedman and Berenson 2017). Changes in extreme BMI over time will be reflected as larger changes in extended z-scores compared to the original z-scores (Cole et al. 2005; Berkey and Colditz 2007; Freedman, Butte, Taveras, Goodman, Ogden, et al. 2017).

In summary, the extended method preserves the use of the current CDC growth charts for the usual CDC LMS-derived calculations of percentiles and z-scores up to the 95th percentile. Extending the reference population above the 95th percentile of the CDC growth charts with more recent national data allows calculation of percentiles and z-scores beyond what was previously recommended and may be useful for monitoring extreme BMI values throughout childhood development in clinical and research settings. As a general methodology, it could also be applied to other body measures in the CDC growth charts.

Acknowledgement

The authors thank Dr. Yulei He for his assistance in checking statistical models/errors and giving suggestions in the draft.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Disclaimer

The findings and conclusions in this report are those of the authors and not necessarily the official position of the Centers for Disease Control and Prevention.

ORCID

David S. Freedman http://orcid.org/0000-0001-8014-0289
Craig M. Hales http://orcid.org/0000-0002-2355-1177

References

Berkey CS, Colditz GA. 2007. Adiposity in adolescents: change in actual BMI works better than change in BMI z score for longitudinal studies. Ann Epidemiol. 17(1):44–50.

CDC National Center for Health Statistics. 2020. NHANES questionnaires, datasets, and related documentation. [accessed Jun 19]. https://wwwn.cdc.gov/nchs/nhanes/Default.aspx.

Centers for Disease Control and Prevention (CDC). 2020a. Percentile data files with LMS values. [accessed Jun 23]. http://www.cdc.gov/growthcharts/percentile_data_files.htm.

Centers for Disease Control and Prevention (CDC). 2020b. A SAS program for the 2000 CDC growth charts. [accessed Jun 23]. https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm.

Cleveland WS. 1979. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 74(368):829–836.

Cole TJ. 1990. The LMS method for constructing normalized growth standards. Eur J Clin Nutr. 44(1):45–60.

Cole TJ. 2010. Extreme percentiles of the 2000 Centers for Disease Control and Prevention BMI chart and the LMS method. Am J Clin Nutr. 91(3):814–814.

Cole TJ, Faith MS, Pietrobelli A, Heo M. 2005. What is the best measure of adiposity change in growing children: BMI, BMI %, BMI z-score or BMI centile? Eur J Clin Nutr. 59(3):419–425.

Cole TJ, Green PJ. 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med. 11(10):1305–1319.

Flegal KM, Cole TJ. 2013. Construction of LMS parameters for the Centers for Disease Control and Prevention 2000 growth charts. Natl Health Stat Rep. 9(63):1–3.

Flegal KM, Wei R, Ogden CL, Freedman DS, Johnson CL, Curtin LR. 2009. 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. 90(5):1314–1320.
Freedman DS, Berenson GS. 2017. Tracking of BMI z scores for severe obesity. Pediatrics. 140(3):e20171072.
Freedman DS, Butte NF, Taveras EM, Goodman AB, Blanck HM. 2017. Longitudinal changes in BMI z-scores among 45 414 2-4-year olds with severe obesity. Ann Hum Biol. 44(8):687–692.
Freedman DS, Butte NF, Taveras EM, Goodman AB, Ogden CL, Blanck HM. 2017. The limitations of transforming very high body mass indexes into z-scores among 8.7 million 2- to 4-year-old children. J Pediatr. 188:50.e1–56.e1.
Freedman DS, Butte NF, Taveras EM, Lundeen EA, Blanck HM, Goodman AB, Ogden CL. 2017. BMI z-scores are a poor indicator of adiposity among 2- to 19-year-olds with very high BMIs, NHANES 1999-2000 to 2013–2014. Obesity (Silver Spring)). 25(4):739–746.
Freedman DS, Woo JG, Ogden CL, Xu JH, Cole TJ. 2019. Distance and percent distance from median BMI as alternatives to BMI z-score. Br J Nutr. 124(5):1–8.
Fryar C, Carroll M, Ogden CL. 2018. Prevalence of overweight and obesity among children and adolescents aged 2–19 years: United States, 1963–1965 through 2013–2014. [accessed 2020 Jun 18]. https://www.cdc.gov/nchs/data/hestat/obesity_child_15_16/obesity_child_15_16.htm.
Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, Wei R, et al. 2002. 2000 CDC growth charts for the United States: methods and development. Vital Health Stat. 11(246):1–190.
Ogden CL, Flegal KM. 2010. Changes in terminology for childhood overweight and obesity. National Health Statistics Reports. 25:1–5.
US Preventive Services Task Force (USPSTF), Grossman DC, Bibbins-Domingo K, Curry SJ, Barry MJ, Davidson KW, Doubeni CA, et al. 2017. Screening for obesity in children and adolescents: US preventive services task force recommendation statement. JAMA. 317(23):2417–2426.
Wei R, Ogden C, Parsons V. 2019. Using a half-normal distribution to track extreme BMI values among youth. Presented at the Joint Statistical Meetings, Section on Statistics in Epidemiology. Denver, 2019. http://www.eventscribe.com/2019/JSM/assets/pdf/1199649.pdf.
Woo JG. 2009. Using body mass index Z-score among severely obese adolescents: a cautionary note. Int J Pediatr Obes. 4(4):405–410.
World Health Organization. Department of Nutrition for Health and Development. 2006. 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. http://www.who.int/childgrowth/standards/technical_report/en/.

Appendix - Estimation of σ_a

The half-normal distribution is specified by the probability density function

$$f(BMI^*) = 2\frac{1}{\sqrt{2\pi}\sigma_a}e^{-(BMI^*)^2/(2\sigma_a^2)}, BMI^* \geq 0,$$

with shape parameter σ_a. $\mathrm{BMI}^* = (\mathrm{BMI}-\mathrm{BMI}_{95th_a})$ where $\mathrm{BMI}_{95^{th}a}$ is the BMI at a given sex- and age-specific 95th percentile at 6-month age interval a. Two important properties of the half-normal distribution are:

$$\mathrm{P1}: E(\mathrm{BMI}^*) = \sqrt{\frac{2}{\pi}}\sigma_a \text{ , and}$$

$$\mathrm{P2}: E(\mathrm{BMI}^{*2}) = \sigma_a^2.$$

Based on relation P1 above, a finite population version of σ_a can be considered as a parameter representing the population mean of a hypothesised finite population. This population parameter has form $\sqrt{\frac{\pi}{2}}\frac{\sum_{i=1}^{N_a}\mathrm{BMI}_{ai}}{N_a}$, where the population has N total units, with class totals N_a and BMI_{ai}^* the individual's excess BMI above the 95th population percentile. The standard complex-survey weighted estimator is

$$\hat{\sigma_a} = \sqrt{\frac{\pi}{2}}\frac{\sum_{i=1}^{n_a} w_i \mathrm{BMI}_i^*}{\sum_{i=1}^{n_a} w_i},$$

where a is the sex-specific 6-month age interval, n_a is the sample size in interval a, and, w_i is the NHANES survey weight for individual i. This functional form is considered as an approximately unbiased estimator for the population mean. Relation P2 leads to an estimator of similar form for σ_a^2, but a required square root may lead to bias. After practical trials, an estimator based on property P1 appeared to provide the best fit with the data.