

# An improved algorithm to harmonize child overweight and obesity prevalence rates

Tim J. Cole<sup>1</sup>  | Tim Lobstein<sup>2,3</sup>

<sup>1</sup>University College London Great Ormond Street Institute of Child Health, London, UK

<sup>2</sup>World Obesity Federation, London, UK

<sup>3</sup>Centre for Health Economics & Policy Innovation, Imperial College, London, UK

## Correspondence

Tim J. Cole, UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.  
Email: [tim.cole@ucl.ac.uk](mailto:tim.cole@ucl.ac.uk)

## Summary

**Background:** Prevalence rates of child overweight and obesity for a group of children vary depending on the BMI reference and cut-off used. Previously we developed an algorithm to convert prevalence rates based on one reference to those based on another.

**Objective:** To improve the algorithm by combining information on overweight and obesity prevalence.

**Methods:** The original algorithm assumed that prevalence according to two different cut-offs A and B differed by a constant amount  $dz$  on the z-score scale. However the results showed that the z-score difference tended to be greater in the upper tail of the distribution and was better represented by  $b \times dz$ , where  $b$  was a constant that varied by group. The improved algorithm uses paired prevalence rates of overweight and obesity to estimate  $b$  for each group. Prevalence based on cut-off A is then transformed to a z-score, adjusted up or down according to  $b \times dz$  and back-transformed, and this predicts prevalence based on cut-off B. The algorithm's performance was tested on 228 groups of children aged 6–17 years from 20 countries.

**Results:** The revised algorithm performed much better than the original. The standard deviation (SD) of residuals, the difference between observed and predicted prevalence, was 0.8% ( $n = 2320$  comparisons), while the SD of the difference between pairs of the original prevalence rates was 4.3%, meaning that the algorithm explained 96.7% of the baseline variance (88.2% with original algorithm).

**Conclusions:** The improved algorithm appears to be effective at harmonizing prevalence rates of child overweight and obesity based on different references.

## KEYWORDS

CDC, harmonization, IOTF, obesity, overweight, prevalence, WHO

## 1 | INTRODUCTION

Child obesity continues to be a major public health concern, and it is important for policy to be able to document its prevalence, across ages and times, in children throughout the world. Two recent large

studies, the NCD Risk Factor Collaboration (NCD-RisC)<sup>1</sup> and the Global Burden of Disease (GBD),<sup>2</sup> have recorded prevalence rates of overweight and obesity by age and sex in 200 countries.

The two studies are important and valuable—but they also highlight a fundamental weakness in that their results are not directly

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. Pediatric Obesity published by John Wiley & Sons Ltd on behalf of World Obesity Federation.

comparable. The NCD-RisC defined child overweight and obesity using the body mass index (BMI) growth standard and reference of the World Health Organization (WHO),<sup>3,4</sup> whereas the GBD used the International Obesity TaskForce (IOTF) reference cut-offs.<sup>5,6</sup> The two references when applied to the same group of children give different prevalence rates of overweight and obesity, and there is no simple way to convert from one to the other. This means that the two sets of results cannot be compared or combined, a *silo effect* which represents a major waste of research effort.

This is an issue of *data harmonization*, and our previous paper<sup>7</sup> introduced an algorithm to address it. The algorithm takes prevalence rates based on one reference, for example, WHO, and estimates what they would be if based on another reference, for example, IOTF. It works by transforming prevalence to the normal equivalent deviate (z-score) scale; this allows switching between references by adjusting the prevalence z-score up or down by an amount that depends only on the two references and the age and sex of the target group of children. The adjusted z-score is then back-transformed to give an estimate of prevalence according to the other reference.

The aim was that the algorithm be both reversible (i.e. predicting from one reference to another and then back again returning the original prevalence) and generalizable (i.e. applying to all datasets). In practice the algorithm worked reasonably well, explaining 88% of the variance.<sup>7</sup> The 12% unexplained error was still relatively large though, making it less than ideal for routine use. We now show that the algorithm can be materially improved. As previously described it focused on individual prevalence rates of overweight or obesity, and ignored the fact that most studies measure both overweight and obesity. Taking the two rates together provides extra information, which reduces the algorithm's unexplained error from 12% to 3%.

The aim of the paper is to describe the improved algorithm and validate it using data on 480 000 children aged 6–17 years from 20 countries.

## 2 | METHODS

### 2.1 | BMI z-score cut-offs

Child overweight and obesity are defined as BMI exceeding an age-sex-specific cut-off that corresponds to a particular BMI z-score. Table 1 gives the z-score cut-offs for overweight and obesity according to the IOTF and WHO references, and also the US Centers for Disease Control and Prevention (CDC) reference.<sup>8</sup> The IOTF z-score cut-offs correspond to sex-specific z-scores of BMI at age 18, 25 kg/m<sup>2</sup> for overweight and 30 kg/m<sup>2</sup> for obesity, for age 2–18.<sup>5</sup> The WHO cut-offs are whole numbers from 1 to 3 that apply to both sexes, with higher cut-offs for age 0–5 than for age 5–19. The CDC cut-offs are the 85th (overweight) and 95th (obesity) centiles of the BMI distribution, for age 2–20 in both sexes.

Note that CDC defines overweight prevalence as excluding obesity prevalence, whereas IOTF and WHO include it. The algorithm requires overweight to include obesity, and where it does not obesity

**TABLE 1** BMI z-score cut-offs for defining overweight and obesity by age and sex according to the WHO, IOTF and CDC references

Reference	Age (years)	Sex	BMI z-score cut-off	
			Overweight	Obesity
WHO	0–5 <sup>a</sup>	Both	2	3
	5–19	Both	1	2
IOTF	2–18	Boys	1.31	2.29
	2–18	Girls	1.24	2.19
CDC	2–20	Both	1.04 to <1.64 <sup>b</sup>	1.64

<sup>a</sup>WHO age 0–5 uses weight-for-height not BMI.

<sup>b</sup>CDC overweight prevalence excludes obesity prevalence.

is added to overweight. Overweight prevalence as reported here always includes obesity prevalence.

### 2.2 | Distributions and QQ plots

The frequency distribution of BMI is markedly skew to the right,<sup>7</sup> and to adjust for it the three BMI references were constructed using versions of the LMS method.<sup>9</sup> This ensured that BMI z-score in the three reference populations was normally distributed. However, when applying the references to other populations normality cannot be assumed—in the decades since the reference data were collected mean BMI has increased and the distribution of BMI has shifted to the right, as seen by increasing rates of overweight and obesity worldwide. This has affected the distribution of BMI z-score in a way that is poorly documented.

A useful way to visualize the frequency distribution of BMI z-score is with the quantile-quantile (QQ) plot, which plots obesity prevalence against the corresponding z-score cut-off. Expressing the prevalence as a z-score (e.g. 5% prevalence equals z-score 1.645) simplifies the plot by plotting z-score against z-score. If the data are normally distributed this gives a straight line QQ plot. And in addition, if the data are from the reference population, that is, standard normal, the z-score prevalence is identical by definition to the z-score cut-off (the one defines the other) and the QQ plot is a straight line with slope 1 and intercept 0.

Prevalence  $p$  is converted to z-score  $Z$  using the formula:  $Z = -\Phi^{-1}(p)$ , where  $\Phi^{-1}$  is the inverse cumulative normal distribution, and the minus sign reflects BMI being above rather than below the cut-off—a higher obesity cut-off gives a lower prevalence. The z-score prevalence is called  $Z$  (upper case) to distinguish it from the z-score cut-off  $z$  (lower case), so the QQ plot consists of  $Z$  plotted against  $z$ . Based on reference data the plot corresponds to the straight line  $Z = z$ . For data from a specified target group the QQ plot is constructed as follows: against each z-score  $z$  is plotted the z-score prevalence  $Z$  of individuals in the group whose BMI z-score matches or exceeds  $z$ .

Figure 1 shows QQ plots for the CDC (green), IOTF (orange) and WHO (blue) references based on their own reference data (single

dotted line, above), and also based on data from a target group of 3699 6-year-old boys from the Health Survey for England (solid lines, below).<sup>10</sup> Prevalence is plotted as  $Z$  on the left axis and as a percentage on the right axis, with the percentage falling as  $Z$  rises. Also shown are the six reference overweight and obesity cut-offs (vertical dotted lines), and on each line there are two points, the upper one corresponding to the reference prevalence and the lower one to the target group prevalence. It is clear that the overweight and obesity rates differ from one reference to another. In addition, comparing the points on each vertical line, overweight is nearly twice as common in the target group compared to the reference populations, and obesity is up to four times as common, showing how prevalence has increased over time.

It is also clear that the three target group QQ plots differ in shape. Focusing on the region covered by the six cut-offs (from  $z=1$  to  $z=2.3$ ) the plots are (a) all close to a straight line; (b) appreciably lower than the reference QQ plots; and (c) shallower in slope for IOTF and WHO (though not CDC). This means that they can be summarized as straight lines by extending the formula  $Z = z$  to  $Z = b \times z + c$ , where the slope  $b$  and intercept  $c$  differ between references and can be estimated by linear regression.

### 2.3 | The original algorithm and the mean z-score cut-off

The original algorithm assumed implicitly that the QQ plot for each target group could be summarized as the line  $Z = z + c$ , that is, with  $b = 1$ .<sup>7</sup> This corresponds to a downward shift in the QQ plot as measured by the intercept  $c$ . However,  $z$  and  $c$  are measured on the reference's z-score scale, so to compare two references a common z-score

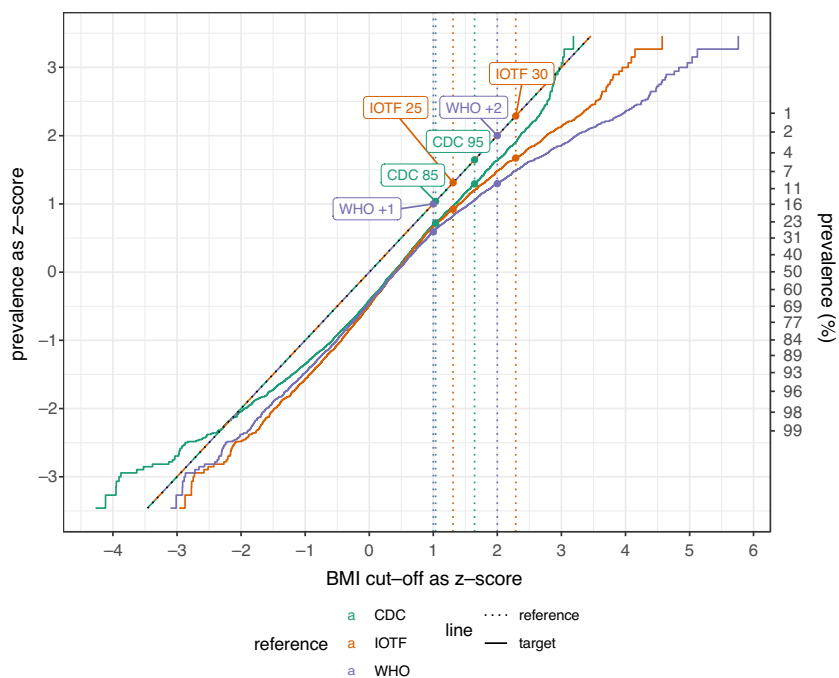
scale is needed. This was achieved as follows: each z-score cut-off was converted to its corresponding BMI cut-off for the group's age and sex (e.g. boys with mean age 6.5 in Figure 1). This BMI was in turn expressed as an age-sex-specific z-score according to the other reference. The two z-scores then relate to the same BMI cut-off, and hence to the same prevalence rate. The average of the two is a z-score on a scale that links to the BMI cut-off and the corresponding prevalence, and which is symmetric in the two references. Effectively it represents an average QQ plot, midway between the QQ plots for the two references. On this common scale the intercept  $c$  is the same for the two references, and their z-score cut-offs are directly comparable.

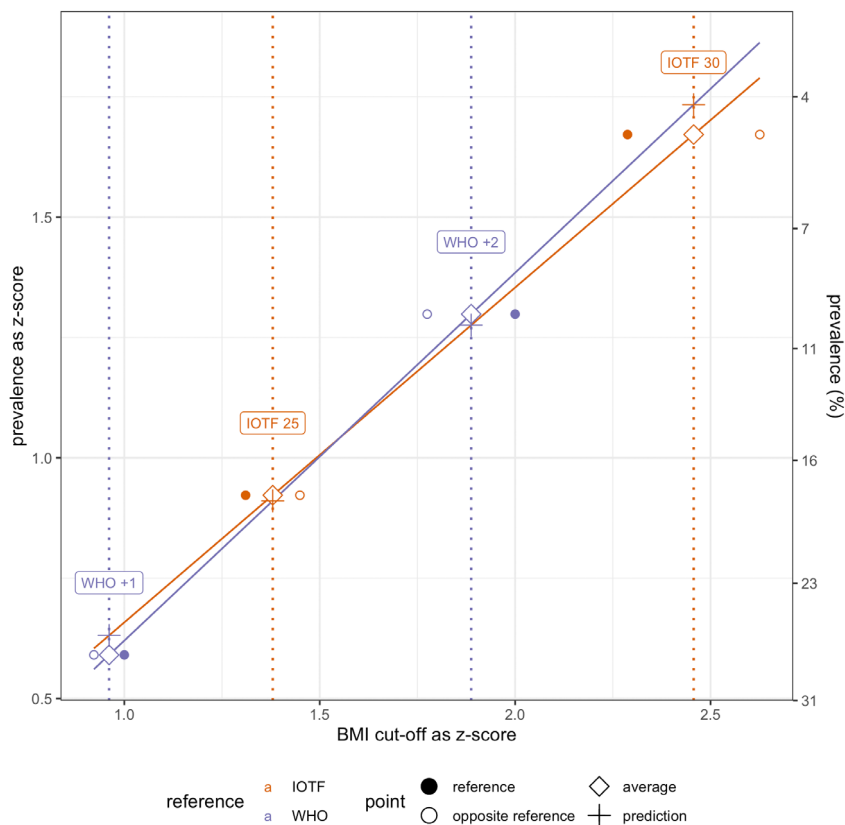
It follows that  $Z_A = z_A + c$  and  $Z_B = z_B + c$ , where the subscripts indicate references A and B. By difference  $dZ = dz$  where  $dZ = Z_B - Z_A$  and  $dz = z_B - z_A$ . This allowed  $Z_B$  to be estimated as  $Z_A + dz$ , or equivalently  $Z_A$  as  $Z_B - dz$ . In other words, the difference in z-score prevalence between the references should match the difference between the corresponding mean z-score cut-offs. But we showed<sup>7</sup> that the formula  $dZ = dz$  was unreliable and the modified formula  $dZ = b \times dz$  worked better, with values of  $b$  per group between 0.7 and 1.0.<sup>7</sup> This is the same as the QQ plot formula  $Z = b \times z + c$ , with slope  $b$  rather than 1.

### 2.4 | Improving the algorithm to estimate $b$

The improved algorithm links overweight and obesity prevalence for each target group to estimate  $b$ —the two points on each QQ plot in Figure 1 define the slope. So a requirement is that target group overweight and obesity prevalence rates should both be available, and in addition they should differ from each other and both be greater than

**FIGURE 1** Quantile-quantile (QQ) plots of BMI z-score in 3699 six-year-old boys from the Health Survey for England according to the CDC, IOTF and WHO references (solid lines). QQ plots based on the reference data are also shown (dotted), a straight line with slope 1 and intercept 0. Vertical dotted lines mark the overweight and obesity cut-offs WHO +1, CDC 85, IOTF 25, CDC 95, WHO +2 and IOTF 30. Prevalence is plotted against BMI z-score cut-off, with prevalence in z-score units on the left axis and as a percentage on the right axis. The points on each vertical cut-off line mark the corresponding prevalence rates for overweight and obesity in the reference population (above) and the target group (below)





**FIGURE 2** Plots of prevalence against z-score cut-off for the IOTF (orange) and WHO (blue) overweight and obesity cut-offs in Figure 1, with prevalence plotted on the z-score scale on the left axis and as a percentage on the right axis. Each prevalence rate is plotted against z-scores for the reference (filled circles), the other reference (open circles) and their average (open diamonds). Regression lines of z-score prevalence on average z-score for WHO and IOTF predict prevalence rates for the other reference (+ signs). See text and Table 2 for details

zero (as zero prevalence corresponds to an infinite z-score). The workflow is as follows:

1. Specify the target group's mean age and sex
2. Obtain the target group prevalence rates of overweight (including obesity) and obesity based on reference A
3. Transform to z-scores  $Z_{ow,A}$  and  $Z_{ob,A}$  where the subscripts indicate *overweight* and *obesity* for reference A
4. Fetch from Table S1, for the target group's mean age and sex, the z-score cut-offs  $z_{ow,A}$  and  $z_{ob,A}$  for reference A and  $z_{ow,B}$  and  $z_{ob,B}$  for reference B
5. Calculate  $b_A = \frac{Z_{ob,A} - Z_{ow,A}}{Z_{ob,A} - Z_{ow,A}}$
6. Calculate  $Z_{ow,B} = Z_{ow,A} + b_A \times (z_{ow,B} - z_{ow,A})$  and  $Z_{ob,B} = Z_{ob,A} + b_A \times (z_{ob,B} - z_{ob,A})$
7. Back-transform  $Z_{ow,B}$  and  $Z_{ob,B}$  to obtain predicted prevalence rates for overweight and obesity based on reference B.

## 2.5 | A worked example

Figure 2 focuses on the QQ plots of Figure 1 for the IOTF and WHO references around the overweight and obesity z-score cut-offs. The cut-offs are shown as vertical dotted lines, each plotted at the relevant average z-score. Each prevalence rate is plotted against z-scores for the reference (filled circles), the other reference (open circles), and their average (open diamonds). Solid lines are drawn through the average z-score diamonds for each reference and extended across the

range. Predicted prevalence rates for each cut-off are then read off the other reference line (+ signs). The observed (diamond) and predicted (+ sign) prevalences are very similar to each other, and the residual errors (where  $residual = observed - predicted$ ) are small whether WHO predicts IOTF or IOTF predicts WHO. The reason for the good fit is that the four diamonds lie close to a straight line, with correlation  $r = 0.9990$  ( $p = 0.0009$ ). This ensures that the two reference lines are very similar in slope.

Table 2 presents the numbers underlying Figure 2, that is, the average z-scores,  $dz$  and  $b$ , and the observed and predicted prevalence rates of overweight and obesity for the target group, as percentages and z-scores. The largest error in prevalence is 1.3%, while the other errors are all 0.6% or less (Table 2).

## 2.6 | Data

To test the algorithm, two datasets were used from the previous paper<sup>7</sup>: the Childhood Obesity Surveillance Initiative (COSI) study by Wijnhoven et al.<sup>11</sup> provided overweight (including obesity) and obesity prevalence based on the IOTF and WHO references, in 225 190 primary school boys and girls aged 6–9 years across 13 European countries during school year 2009/10. The study provided data for 52 distinct country-age-sex groups, with age grouped to the last completed year. Deren et al.<sup>12</sup> published prevalence rates of overweight (net of obesity) and obesity based on IOTF, WHO, and CDC cut-offs in 18 144 Ukrainian children and adolescents aged

**TABLE 2** Summary statistics for WHO and IOTF overweight and obesity prevalence and z-score cut-offs in boys aged 6, for the two reference populations and a target group of 3699 boys from the Health Survey for England<sup>10</sup>

	Reference cut-off	WHO +1 overweight	IOTF 25 overweight	WHO +2 obesity	IOTF 30 obesity
Reference	Z-score cut-off	1.00 <sup>a</sup>	1.31 <sup>b</sup>	2.00 <sup>a</sup>	2.29 <sup>b</sup>
	BMI cut-off (kg/m <sup>2</sup> )	16.9	17.7	18.7	20.1
	Z-score cut-off with other reference	0.92	1.45	1.77	2.63
	Average z-score cut-off	0.96 <sup>c</sup>	1.38 <sup>c</sup>	1.89 <sup>d</sup>	2.46 <sup>d</sup>
	Z-score difference dz	+0.42 <sup>c</sup>	-0.42 <sup>c</sup>	+0.57 <sup>d</sup>	-0.57 <sup>d</sup>
Target group	Observed prevalence (z-score)	27.7% (0.59) <sup>a</sup>	17.8% (0.92) <sup>b</sup>	9.7% (1.30) <sup>a</sup>	4.7% (1.67) <sup>b</sup>
	Slope <i>b</i>	0.76 <sup>a</sup>	0.70 <sup>b</sup>	0.76 <sup>a</sup>	0.70 <sup>b</sup>
	Prevalence predicted by IOTF (z-score) <sup>e</sup>	26.4% (0.63)	–	10.1% (1.28)	–
	Prevalence predicted by WHO (z-score) <sup>e</sup>	–	18.1% (0.91)	–	4.1% (1.73)
	Observed minus predicted prevalence (z-score)	+1.3% (-0.04)	-0.3% (+0.01)	-0.4% (+0.02)	+0.6% (-0.06)

<sup>a</sup> $b_{WHO} = \frac{1.30 - 0.59}{2.00 - 1.00} = 0.76.$

<sup>b</sup> $b_{IOTF} = \frac{1.67 - 0.92}{2.29 - 1.31} = 0.70.$

<sup>c</sup>Overweight dz = 1.38 - 0.96 = 0.42 (+ve for WHO, -ve for IOTF).

<sup>d</sup>Obesity dz = 2.46 - 1.89 = 0.57 (+ve for WHO, -ve for IOTF).

<sup>e</sup>Calculated as (z-score of prevalence + *b* × dz) for other reference.

6.5–17.5 years, in 22 distinct sex-year groups with age grouped to the nearest year. For the analysis, obesity prevalence was added to overweight prevalence.

In addition, individual data for 231 218 boys and girls aged 5–18 from the following publicly available surveys were obtained (with the kind help of the NCD Risk Factor Collaboration):

- China Health and Nutrition Survey (CHNS): 1991–2015<sup>13</sup> (*n* = 17 975).
- Mexico National Survey of Health and Nutrition (ENSANUT) and Mexican Family Life Survey (ENNViH): 2002–2016<sup>14,15</sup> (*n* = 77 354).
- US National Health And Nutrition Examination Survey (NHANES): 1973–2018<sup>16</sup> (*n* = 39 431).
- UK Health Survey for England (HSE): 1994–2019<sup>10</sup> (*n* = 69 466).
- Russia Longitudinal Monitoring Survey (RLMS): 1993–2005<sup>17</sup> (*n* = 26 992).

Further details of the datasets are in the appendix to the NCD-RisC report.<sup>18</sup> The individual data consisted of sex, age in completed years, and measured height and weight recorded to one decimal place. Individual data for 7157 Algerian children aged 6–18 were also kindly made available,<sup>19</sup> consisting of sex, decimal age, and measured height and weight to one decimal place. For these individual data, body mass index was calculated as  $weight(kg)/height(m)^2$ , and country-age-sex-specific prevalence rates for overweight (including obesity) and obesity by CDC, IOTF and WHO were calculated for each completed year of age in each country, with 26 groups for China, Mexico, USA, UK and Russia, and 24 for Algeria. For the multi-year surveys, data were pooled across years to ensure sufficient numbers in each group and minimize sampling error. For analysis, all group mean ages were rounded to the nearest 0.5 years.

## 2.7 | Statistical analysis

The analysis involves two distinct steps: first estimate *b*, then convert the prevalence rates. However it can be done in a single step using weighted linear regression of *Z* on *z* across the four cut-offs. Two of the prevalences are known and the other two are unknown, so setting weights of 1 for the known prevalences and 0 for the unknown prevalences ensures that *b* is estimated from the known prevalences; however, it also provides fitted values for the unknown prevalences based on their mean z-scores. So in Table 2, IOTF is estimated from WHO using weights (1,0,1,0) (with the cut-offs ranked as in the table) while to estimate WHO from IOTF the weights are (0,1,0,1). With just two known prevalences the regression estimates the intercept and slope with no degrees of freedom for error.

In addition unweighted linear regression including all four cut-offs is used to estimate the linear regression coefficient *b* of *Z* on *z*, the quadratic regression of *Z* on *z*, and the correlation  $r = cor(Z, z)$ . Here there are one or two degrees of freedom for error.

To simplify the calculations, Table S1 provides mean z-scores for the four cut-offs by sex and mean age in half-years from 2 to 18, for the pairs of references WHO-IOTF, IOTF-CDC and CDC-WHO. Values closest in age to the mean age of each target group should be used in the analysis. Supplementary Box S1 provides sample R code for analysing the data.

Altogether the data provided 228 country-sex-age groups, each with prevalence rates based on all three (or for the 52 COSI study groups just WHO and IOTF) references. The reference comparisons for WHO-IOTF, IOTF-CDC and CDC-WHO were fitted simultaneously, by “stacking” the data into a single data frame with  $52 + (228 - 52) \times 3 = 580$  rows, where each row provided four estimates of overweight and obesity prevalence, giving 2320 predictions

**TABLE 3** Summary statistics for predicting overweight and obesity prevalence according to the WHO, IOTF and CDC references, for groups of children indexed by sex, age, country, predicting and predicted reference. The distribution of residuals (where residual = observed – predicted prevalence) is shown in both % and z-score units

Category	Number of groups	Number of children per group (median)	Predicting reference	Predicted reference	Mean observed prevalence (%)	Mean residual prevalence (%)	SD of residual prevalence (%)	SD of residual prevalence (z-score)
Overall	2320	1112			14.4	0.0	0.8	0.05
Overweight	1160	1112			21.2	0.0	1.0	0.04
Obesity	1160	1112			7.7	0.0	0.6	0.06
Overweight	176	1098	CDC	IOTF	18.6	0.0	0.4	0.03
Overweight	176	1098	IOTF	CDC	19.2	0.1	0.6	0.03
Overweight	228	1158	IOTF	WHO	25.6	0.0	1.6	0.06
Overweight	228	1158	WHO	IOTF	20.0	–0.1	0.7	0.03
Overweight	176	1098	WHO	CDC	19.2	0.0	0.6	0.04
Overweight	176	1098	CDC	WHO	23.6	0.0	1.1	0.05
Obesity	176	1098	CDC	IOTF	5.6	–0.1	0.6	0.07
Obesity	176	1098	IOTF	CDC	8.1	0.2	0.5	0.05
Obesity	228	1158	IOTF	WHO	9.5	0.1	0.6	0.05
Obesity	228	1158	WHO	IOTF	6.1	–0.1	0.7	0.08
Obesity	176	1098	WHO	CDC	8.1	–0.1	0.3	0.03
Obesity	176	1098	CDC	WHO	8.5	0.1	0.3	0.03

overall. The percentage of the variance of prevalence explained by the algorithm compared to the baseline model, where prevalence  $p_A$  is predicted by  $p_B$  and vice versa, was calculated as  $100 \left( 1 - \frac{SD(p_A - E(p_A))}{SD(p_A - p_B)} \right)^2$  based on all 2320 points.

The calculations were done in R (version 4.1.3) running in RStudio. A function `ob_convertr2` was written to do the conversion, which is available in the first author's CRAN `sitar` package (version 1.3.0).<sup>20</sup>

### 3 | RESULTS

Data on 481 709 children aged 5–18 from 20 countries were grouped into 228 one-year age groups by country and sex, with median 1112 children per group (minimum 145, lower quartile 737, upper quartile 2483, maximum 26 542).

Within each group, z-score prevalence  $Z$  and mean z-score cut-off  $z$  for the four cut-off points were highly correlated—across the 580 age-sex-country-reference comparisons, the median correlation  $r$  was 0.9994 ( $p = 0.0006$ ), that is, almost perfectly collinear, with only three correlations less than 0.98 and 19 (3%) less than 0.99. The regression coefficient  $b$  based on the four points ranged from 0.4 to 1.5 with mean 0.81 (SD 0.17). Fifty-nine (10%) of the 580 quadratic regressions had a significant quadratic term ( $p < 0.05$ ), 24 positive and 35 negative, indicating some curvature though no material bias.

Table 3 summarizes the distribution of residuals based on the algorithm, where the residual is the difference between observed and predicted prevalence, measured in both percentage and z-score units. The mean residual was close to zero throughout, so the algorithm was

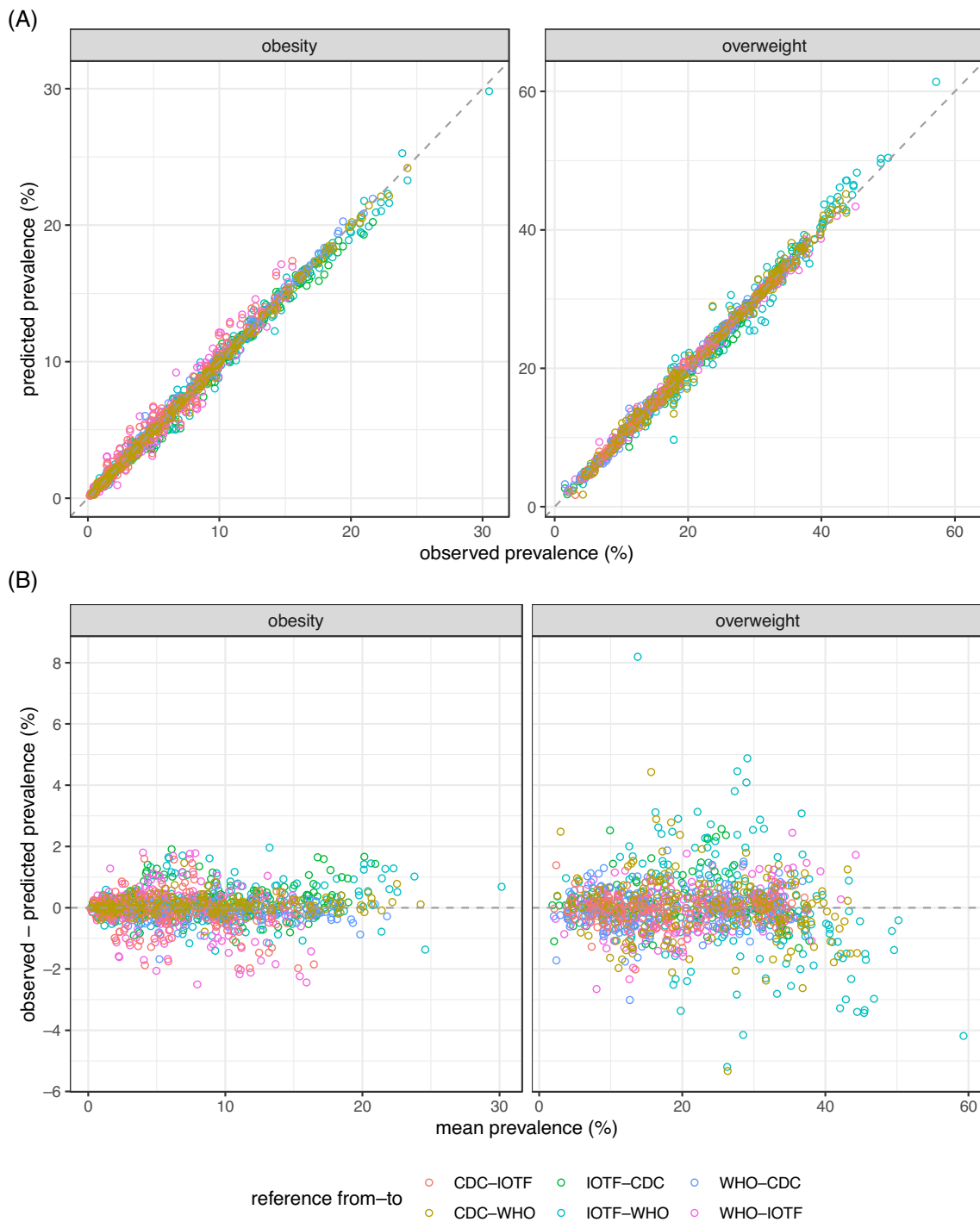
unbiased. The SD of residuals was also small, 0.8% or 0.05 z-score units overall, based on 2320 predictions. For comparison the SD of the difference between prevalence rates  $SD(p_A - p_B)$  was 4.3%, meaning that the algorithm explained 96.7% of the baseline variance in prevalence.

The residual SD was larger for overweight (1.0%) than for obesity (0.6%), though overweight was nearly three times as common as obesity (mean 21.2% versus 7.7%) so in proportional terms its residual was smaller. On the z-score scale the SDs for overweight and obesity were 0.04 and 0.06. Comparing the different pairs of references, the residual SD ranged from 0.4% to 1.6% for overweight and from 0.2% to 0.7% for obesity, with generally larger SDs for higher mean overweight prevalence. On the z-score scale residual SDs ranged from 0.03 to 0.07, unrelated to mean prevalence.

Figure 3 shows the relationship between observed and predicted prevalence, for obesity and overweight separately: as scatterplots (Figure 3A) and as Bland–Altman plots (Figure 3B). The points are colour-coded by predicting and predicted references. Figure 3A shows a very close association between observed and predicted prevalence, with correlations of 0.995 and 0.996 (both  $p < 0.0001$ ) respectively for obesity and overweight, each based on 1160 points.

Figure 3B shows Bland–Altman plots, that is, residual prevalence plotted against the mean of observed and predicted prevalence for overweight and obesity separately. There is no obvious trend in residual versus mean, though variability is less at low prevalence. Four-fifths of the obesity residuals are less than 0.6% in absolute value, maximum 2.5%, while four-fifths of the overweight residuals are less than 1.0%, maximum 8.2%. The largest three overweight residuals are in small groups with 224 or fewer children.





**FIGURE 3** The close relationship between observed and predicted prevalence, shown for obesity and overweight separately: (a) scatterplots and (b) Bland-Altman plots. The 2320 points are colour-coded by the predicting and predicted references. The scatterplots show correlations between observed and predicted prevalence exceeding 0.994. The Bland-Altman plots relate residual (i.e. observed minus predicted) prevalence to mean prevalence, showing that most residuals are smaller than 1% and the scatter is less at low prevalence

#### 4 | DISCUSSION

The findings show that prevalence rates of child overweight and obesity based on one BMI reference, for example, WHO, can to high accuracy be converted to equivalent rates based on another reference, for example, IOTF. Figure 3 demonstrates the tight correlation

of 0.995 between predicted and observed prevalence across 2320 comparisons, where the SD of the residuals is 1.0% for overweight and 0.6% for obesity. This improved algorithm explains 96.7% of the baseline variance in prevalence, as compared to 88.2% achieved by the original algorithm.<sup>7</sup> So improving the algorithm has reduced the residual variance by over four-fifths, from 11.8% to 3.3%. To apply

the algorithm, the only requirements are that the target groups have prevalence rates available for both overweight and obesity, and that the rates are non-zero and different from each other.

The reason why the algorithm fits so well is because overweight and obesity prevalence for the two references, plotted against the corresponding cut-offs, lie close to a straight line, as seen in Figure 1. Based on 228 target group comparisons the median correlation across the four points is 0.9994, and all but three of the 228 correlations exceed 0.98, so the collinearity is quite general. It means that the two target lines in Figure 2 are very similar in slope, so each can be used to predict the other.

This collinearity arises because the BMI distribution QQ plots for each target group retain the linearity of the original reference QQ plots. The QQ plot shows the prevalence plotted against the corresponding z-score cut-off across the z-score range. The dotted line (Figure 1, top), with slope 1 and intercept 0, shows how BMI was distributed in the reference populations. But with the subsequent rise in child obesity the BMI distribution has altered in two ways:

- The whole distribution has shifted to the right; overweight and obesity prevalence have increased, and the QQ plot intercept has shifted downwards (as higher prevalence corresponds to a lower z-score).
- In addition the distribution has generally become more skew, with a relatively heavier upper tail, and this has caused the QQ plot to fall more at the upper than the lower end of the range, making the line shallower.

This means that the target group QQ plot is still close to linear, but its slope and intercept differ from 1 and 0 in a way that varies by country, age and sex. The algorithm estimates the target slope and intercept from the overweight and obesity prevalence rates based on one reference, and converts the rates to the other reference assuming that the same slope and intercept apply—this holds as the four points are collinear. Thus the algorithm is close to reversible—converting from A to B and back again gives prevalence rates very similar to the original values—and it is generalizable in that each target group calibrates itself.

It is this use of overweight and obesity prevalence in tandem that distinguishes the improved algorithm from the original.<sup>7</sup> There each prevalence rate was treated independently, which meant that with just the one prevalence rate the intercept of the relevant QQ plot could be estimated, but not the slope. For this reason the improved algorithm cannot be applied to thinness, as there is just one thinness cut-off per reference not two; however if studies were to report prevalence rates of two or more grades of IOTF thinness, this information could be used in the same way to improve the estimate of WHO or CDC thinness prevalence.

The shapes of the target group QQ plots in Figure 1 are instructive. Those for IOTF and WHO are essentially linear with slope 1 below the WHO +1 cut-off, indicating a similar increase in z-score prevalence over time for underweight and normal weight individuals. However, the plots tend to curve downwards above WHO +1 (the

84th centile) reflecting a heavier upper tail to the BMI distribution, meaning that the z-score prevalence has increased more for those overweight or obese. The CDC plot in contrast curves upwards in both tails of the distribution, which is an indication of left skewness,<sup>21</sup> and this relates to the way the CDC reference was constructed, restricted to data from the 3rd to the 97th centile, that is, from  $-1.9$  to  $1.9$  z-scores.<sup>22</sup> The degree of skewness depends on the relative lengths of the two tails of the distribution, so if the tails are excluded from the analysis the skewness cannot be estimated accurately, and this is what happened with CDC.

The analysis has focused on the IOTF, WHO and CDC references, but it is worth pointing out that the same analysis can be applied to any BMI reference that expresses BMI as a z-score (e.g. based on the LMS method<sup>9</sup>) and which uses z-scores as cut-offs. Also, the same analysis can be applied to a single reference where prevalence is known for two cut-offs, to estimate prevalence at a third cut-off. In this simpler case the z-score cut-offs do not need averaging as there is only the one reference.

The study aimed to develop a reversible and generalizable algorithm, and a strength of the study is that it has been successful as applied to a wide variety of data sources. First, the data cover a wide spectrum of overweight and obesity prevalence, from near 0% up to 60% overweight and 30% obesity (Figure 3). The fact that the fit is adequate for these data implies similar goodness of fit for other studies covering the same prevalence spectrum. Second, having paired prevalence rates for overweight and obesity means that the cut-offs for the known and unknown rates are interleaved—in Figure 2 IOTF overweight lies between WHO overweight and obesity, while WHO obesity lies between IOTF overweight and obesity. These two prevalence rates are estimated by interpolation rather than extrapolation, which increase their precision. Table 3 shows that WHO predicts IOTF overweight more precisely than IOTF predicts WHO overweight, and similarly for WHO-CDC, while for obesity the reverse is true. It is reassuring that the interpolated prevalence rates are more precise than the extrapolated rates, but also reassuring that even the extrapolated rates are relatively precise.

The three references cover different age ranges: 0–19 years for WHO, 2–18 for IOTF and 2–20 for CDC (Table 1). The algorithm was validated here with data from age 5 to 18 years, and it is a limitation of the study that no children under age 5 were included. A complication is that WHO uses weight-for-height rather than BMI under age 5, and BMI differs from weight-for-height in that it adjusts for age as well as height.<sup>23</sup> This suggests that the algorithm may be largely irrelevant under age 5. However for age 2–5 BMI and weight-for-height are very similar:  $BMI = \text{weight}/\text{height}^2$  while  $\text{weight-for-height} = \text{weight}/\text{height}^p$  where  $p$  lies between 1.5 and 2.<sup>23</sup> For this reason, weight-for-height z-score and BMI z-score are likely to be similar for age 2–5, which suggests that converting between WHO weight-for-height prevalence and IOTF or CDC BMI prevalence may be justified. In addition some countries already use the WHO cut-offs for BMI rather than weight-for-height, for example, Canada,<sup>24</sup> and so the algorithm has a role to play there.

A second limitation of the algorithm is that it requires the prevalence rates of overweight and obesity to be non-zero and different



from each other. This means that groups need to be large enough to include several cases of obesity, which implies even more cases of overweight (since overweight prevalence includes obesity prevalence). The groups used here were deliberately made large to minimize sampling error, so that the observed residuals can be attributed to the algorithm rather than to small numbers of cases.

In conclusion, the study has shown that the improved algorithm using paired prevalence rates of overweight and obesity based on one BMI reference is effective at predicting the corresponding prevalence rates based on another reference. A table is provided to simplify the analysis for groups of children using the IOTF, WHO or CDC references.

## ACKNOWLEDGEMENTS

TJC designed the study, did the data analysis, generated the figures and wrote the first draft of the paper. TJC and TL edited the paper and had final approval of the submitted version. We are grateful to Marie Françoise Rolland-Cachera and Majid Ezzati for their helpful comments on earlier drafts of the paper. We would also like thank Majid Ezzati and Bin Zhou very much for their providing access to publicly available data from the NCD Risk Factor Collaboration, and Nora Bahchachi for making available her Algerian data.<sup>19</sup> The following is included as a condition of using the data. This research uses data from China Health and Nutrition Survey (CHNS). We are grateful to research grant funding from the National Institute for Health (NIH), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) for R01 HD30880, National Institute on Aging (NIA) for R01 AG065357, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) for R01DK104371 and R01HL108427, the NIH Fogarty grant D43 TW009077 since 1989, and the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009, Chinese National Human Genome Center at Shanghai since 2009, and Beijing Municipal Center for Disease Prevention and Control since 2011. We thank the National Institute for Nutrition and Health, China Center for Disease Control and Prevention, Beijing Municipal Center for Disease Control and Prevention, and the Chinese National Human Genome Center at Shanghai.

## CONFLICT OF INTEREST

TJC declares the following conflicts of interest: he developed the LMS method with Peter Green,<sup>9</sup> and was first author on papers describing the IOTF cut-offs.<sup>5,6,25</sup> TL was also an author on one of the IOTF papers<sup>6</sup> and was an employee of the IOTF from 2004 to 2014.

## ORCID

Tim J. Cole  <https://orcid.org/0000-0001-5711-8200>

## REFERENCES

- Murray CJL, Aravkin AY, Zheng P, et al. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet*. 2020;396:1223-1249.
- Ezzati M, Bentham J, Di Cesare M, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet*. 2017;390:2627-2642.
- de Onis M, Garza C, Onyango AW, Martorell R. WHO child growth standards. *Acta Paediatr*. 2006;95(Suppl 450):3-101.
- de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull WHO*. 2007;85:660-667.
- Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*. 2000;320:1240-1243.
- Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatr Obes*. 2012;7:284-294.
- Cole TJ, Lobstein T. Exploring an algorithm to harmonize international obesity task Force and World Health Organization child overweight and obesity prevalence rates. *Pediatr Obes*. 2022; In press.
- Must A, Dallal GE, Dietz WH. Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht<sup>2</sup>) and triceps skinfold thickness. *Am J Clin Nutr*. 1991;53:839-846.
- Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med*. 1992;11:1305-1319.
- Anonymous. Health Survey for England. 2022 <https://digital.nhs.uk/data-and-information/areas-of-interest/public-health/health-survey-for-england—health-social-care-and-lifestyles#survey-dataset>
- Wijnhoven TMA, van Raaij JMA, Spinelli A, et al. WHO European childhood obesity surveillance initiative: body mass index and level of overweight among 6-9-year-old children from school year 2007/2008 to school year 2009/2010. *BMC Public Health*. 2014;14:1-16.
- Deren K, Wyszynska J, Nyankovskyy S, et al. Assessment of body mass index in a pediatric population aged 7-17 from Ukraine according to various international criteria—a cross-sectional study. *PLoS ONE*. 2020;15:e0244300.
- Anonymous. China Health and Nutrition Survey. 2022 <https://www.cpc.unc.edu/projects/china>
- Anonymous. Mexico National Survey of Health and Nutrition. 2022 <https://ghdx.healthdata.org/organizations/national-institute-public-health-mexico>
- Anonymous. Encuesta Nacional sobre Niveles de Vida de los Hogares. 2022 <http://www.ennvih-mxfls.org/>
- Anonymous. National Health and Nutrition Examination Survey. <https://www.cdc.gov/nchs/nhanes/index.htm>
- Anonymous. Russia Longitudinal Monitoring Survey. 2022 <https://dataverse.unc.edu/dataset.xhtml?persistentId=doi:10.15139/S3/12438>
- Rodriguez-Martinez A, Zhou B, Sophiea MK, et al. Height and body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries and territories: a pooled analysis of 2181 population-based studies with 65 million participants. *Lancet*. 2020;396:1511-1524.
- Bahchachi N, Dahel-Mekhancha CC, Rolland-Cachera MF, Roelants M, Hauspie R, Nezzal L. Body mass index charts of Algerian children and adolescents (6-18 years). *Archiv Pediatr*. 2017;24:1205-1213.
- Cole T. Super imposition by translation and rotation growth curve analysis R package sitar. Comprehensive R archive Network; 2021.
- van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Stat Med*. 2001;20:1259-1277.
- Flegal KM, Cole TJ. *Construction of LMS Parameters for the Centers for Disease Control and Prevention 2000 Growth Charts*. National Center for Health Statistics; 2013.
- Cole TJ. A critique of the NCHS weight for height standard. *Hum Biol*. 1985;57:183-196.
- Rao DP, Kropac E, Do MT, Roberts KC, Jayaraman GC. Childhood overweight and obesity trends in Canada. *Health Promot Chronic Dis Preven Canada-Res Policy Pract*. 2016;36:194-198.

25. Cole TJ, Flegal KM, Nicholls D, Jackson AA. Body mass index cut offs to define thinness in children and adolescents: international survey. *BMJ*. 2007;335:194-197.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Cole TJ, Lobstein T. An improved algorithm to harmonize child overweight and obesity prevalence rates. *Pediatric Obesity*. 2022;e12970. doi:[10.1111/ijpo.12970](https://doi.org/10.1111/ijpo.12970)