The perils of standardizing infant weight to assess weight change differences across exposure groups

Ann Von Holle [a, *], Kari E. North, PhD [a], Ran Tao, PhD [b, c], Sheila Gahagan, MD, MPH [d, e]

[a] University North Carolina, Chapel Hill, Gillings School of Global Public Health, 8 Department of Epidemiology, Chapel Hill, NC 27599
[b] Department of Biostatistics, University of North Carolina, Chapel Hill, Gillings School of Global Public Health, Chapel Hill, NC 27599
[c] Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN 37203
[d] Department of Pediatrics, School of Medicine, University of California, San Diego, La Jolla, CA 92093
[e] Division of Child Development and Community Health, University of California, San Diego, San Diego, CA 92103

ARTICLE INFO

Article history:
Received 31 January 2018
Accepted 8 April 2018
Available online 25 April 2018

Keywords:
Infant
Growth
Body size
Body weight
Reference standards

ABSTRACT

Purpose: When conducting analyses of child weight growth trajectories, researchers commonly use Z-scores from a standard instead of the observed weights. However, these Z-scores, calculated from cross-sectional data, may introduce methodological limitations when used in the context of longitudinal analyses. We assessed analytic limitations when analyzing infant growth data with three anthropometric measures: weight and the corresponding Z-scores and percentiles from a standard.
Methods: We undertook a series of Monte Carlo simulations and compared tests of differences in postnatal weight change across time (growth velocity) between two exposure groups. Models with the observed weight outcome were compared to the corresponding weight World Health Organization (WHO) Z-score or weight percentile outcomes. We calculated power, type I error, and median product term coefficient estimates to assess differences between the models.
Results: There was lower power to detect velocity differences across exposure groups for WHO Z-scores and percentiles as outcomes compared to the use of observed weight values. We also noted instances in which velocity differences between exposed and unexposed groups were in the opposite direction in analyses with WHO Z-score outcomes.
Conclusions: In our simulations of infant weight velocity differences across exposure groups, we observed lower power and effect inconsistencies when applying a standard-derived Z-score transformation. These results emphasize the need for careful consideration of the appropriate scale when assessing infant growth trajectories across categorical groups.

Introduction

Analyses of growth trajectories are expanding in tandem with the growing interest in life course epidemiology and the increased availability of longitudinal data collected during the postnatal period. When modeling growth, an investigator may consider using data transformations, in line with standards that have been developed in the field for cross-sectional assessment of weight in childhood. For example, Z-scores or percentiles derived from a standard, such as the World Health Organization (WHO), or derived from a reference, such as from the Centers for Disease Control, can be implemented. Z-scores for infant growth (up to the age of 2 years), derived from the 2006 WHO growth standards [1], provide a measure of relative position on the distributional scale of optimal infant growth following a standard normal distribution [2]. A Z-score value of zero represents the mean of the optimal distribution for a respective age and sex distribution. A Z-score unit change represents the change in one standard deviation of the population standard distribution for the respective anthropometric measure. Percentiles map to Z-scores, and they provide a measure bounded by the lowest value of 0 and highest value of 100, which represents the percentage of the distribution of the standard population below this value.

Using Z-scores from a standard such as WHO or Centers for Disease Control can have advantages in cross-sectional analyses, such as providing linear sex- and age- independent measures of weight outcomes for individuals [2] and the ability to compare estimates across different studies. However, the use of Z-scores in longitudinal analyses may be inappropriate, given their derivation

* Corresponding author. University of North Carolina - Chapel Hill, Chapel Hill, NC 27599 7435.
E-mail address: vonholle@email.unc.edu (A. Von Holle).

https://doi.org/10.1016/j.annepidem.2018.04.006

from cross-sectional populations. Despite this concern, longitudinal analyses commonly use Z-score transformations.

Other concerns related to the use of Z-score transformations in longitudinal infant weight trajectories have not been widely explored but may include diminished power, a liberal type I error rate, and biased estimates of group differences. For example, a previous study of changes in growth during adolescence [3] demonstrated diminished power to detect effects when using Z-scores compared to actual body mass index. To date, no such studies comparing various data transformations in the study of group differences in weight change during infancy have been conducted.

Our primary aim in this article was to assess the analytic limitations of using Z-scores from a standard in place of observed weight values when conducting longitudinal analyses of infant growth. We used Monte Carlo data simulations to assess differences in power, type I error, and effect estimation using longitudinal data across three anthropometric measures: weight, weight WHO Z-scores, and weight percentiles. These results may provide guidance in choosing an appropriate outcome measure when evaluating change in weight across groups in infancy.

Materials and methods

To assess differences in power, type I error, and bias when using different outcome measures, we followed a two-step process. First, we simulated infant weight growth data using parameters from published literature to provide individual infant growth curve samples from actual populations. These data supplied the observed values in the following step. Second, we estimated weight change differences in the simulated data across two time points with analysis of variance (ANOVA) models. Comparisons were made for three outcomes including the original weight value, a WHO Z-score, and matching percentile.

Growth data simulation

Simulated infant weight growth data were based on a Reed first order parametric model [4,5], a typical form being

$$y = \beta_{0i} + \beta_{1i} \cdot t + \beta_2 \cdot \ln(t) + \frac{\beta_3}{t}$$

We use a time transformation suggested for this model, $t^* = \frac{t+9}{9}$ [4], to allow a time point (t) of birth at 0 months. The Reed first order model has several advantages, one of which includes linearity in parameter terms and thus simplicity in model fit —an advantage when conducting a large number of iterations in a simulation. This model has also been shown to have a good model fit when compared to other nonlinear models such as the Jenns-Bayley model [5]. Parameters specified in the simulations were drawn from publicly accessible estimates in the literature for three different countries [6]. The three countries included in the analyses were Italy, Portugal, and Chile. This variety of populations may provide a diverse and realistic assessment of model performance when comparing the weight values as outcome measures versus WHO Z-scores or related measures such as percentiles.

To accommodate exposure group differences in the Reed first order growth model, we added two terms, β_4 and β_5 to create two distinct growth shapes according to the binary exposure levels:

$$y_{ij} = \beta_{0i} + \beta_{1i} \cdot t^* + \beta_{2i} \cdot \ln(t^*) + \frac{\beta_3 i}{t^*} + \beta_4 \cdot group + \beta_5 \cdot t^* \cdot group + e_{ij}$$

The sum of the β_4 and β_5 terms represented a difference of weight at baseline across exposure groups. Furthermore, the β_5 term represented a difference in weight change over time between the two groups.

In the model, y_{ij} represents the weight in kilograms for each subject (i) at seven time points (j) evenly spaced from 0 to 6 months. The group variable represents what one would consider an exposure variable in an analysis with one index category (exposed) and one referent (unexposed) and was split at random using a binomial distribution with probabilities of 0.2 or 0.5. The referent, considered to be the unexposed group, matched the Reed parameters from the growth model samples obtained from the literature as mentioned previously [6]. The total sample size was 1000 people.

The models included random effects for both the intercept, β_0, and each of the three time terms, β_1, β_2, and β_3. Residuals, e_{ij}, were assumed to have SD values of $\sigma = 0.5$. Missing data were generated with zero to one missing time point per person out of the seven total, creating potential for unevenly spaced time points for each person.

To examine any differences in weight change estimates, given the standardization status of the outcome measure, we chose a two by two factorial design between baseline weight status and growth status across exposure groups. Baseline weight status was either no difference or a difference in baseline average weight matching direction of slope difference between the exposure groups. Slope differences between the exposure groups were either positive or negative. It should be noted that due to the adapted Reed model, baseline differences were dependent on the sum of β_4 and β_5. The first combination, $\beta_4 = 0$ and $\beta_5 = 0.5$, resulted in the index group with a higher baseline weight at t = 0 months ($t^* = 1$) ($\beta_4 + \beta_5 = 0 + 0.5 = 0.5$) and faster weight change ($\beta_5 = 0.5$). The second combination, $\beta_4 = -0.5$ and $\beta_5 = 0.5$, resulted in the index group with an equal baseline weight ($\beta_4 + \beta_5 = -0.5 + 0.5 = 0$) and faster weight change ($\beta_5 = 0.5$). The third combination, $\beta_4 = 0.5$ and $\beta_5 = -0.5$, resulted in the index group with an equal baseline weight ($\beta_4+\beta_5 = 0.5 - 0.5 = 0$) and slower weight change ($\beta_5 = -0.5$). Finally, the fourth combination, $\beta_4 = 0$ and $\beta_5 = -0.5$, resulted in the index group with a lower baseline weight ($\beta_4+\beta_5 = 0 - 0.5 = -0.5$) and slower weight change ($\beta_5 = -0.5$). The four groups from the two by two factorial design will be referred to as "baseline+ and slope+," "baseline= and slope+," "baseline= and slope−," and "baseline− and slope−," respectively.

For the simulated weight growth data, we chose the β_4 and β_5 values after visually examining plots so they reflected biologically plausible differences in baseline weight and growth (see Fig. 1). For the two scenarios with a baseline weight difference in the exposed group relative to the referent group, the difference appeared to be around 25 percentile units. As a sensitivity analysis, we examined two more sets of values including combinations of 0.1 and 0.2 for β_4 and β_5 to determine if our findings were consistent with smaller differences across exposure groups. Following simulation of the weight growth from the model above, we calculated their corresponding Z-scores using the WHO growth standards [7]. Percentiles were derived from the Z-scores using the cumulative probability for the standard normal distribution.

Growth data analysis

Common practice includes comparison of mean weights across two time points using ANOVA. We chose that model to estimate weight change from birth to 6 months for its simplicity and freedom from assumptions regarding shape of weight trajectories over time. The two time groups are represented by a dichotomous variable, month6, and exposure also represented by a dichotomous exposure values: "exposed" and "unexposed" groups. The product term β_3 covers the difference in weight change across exposure

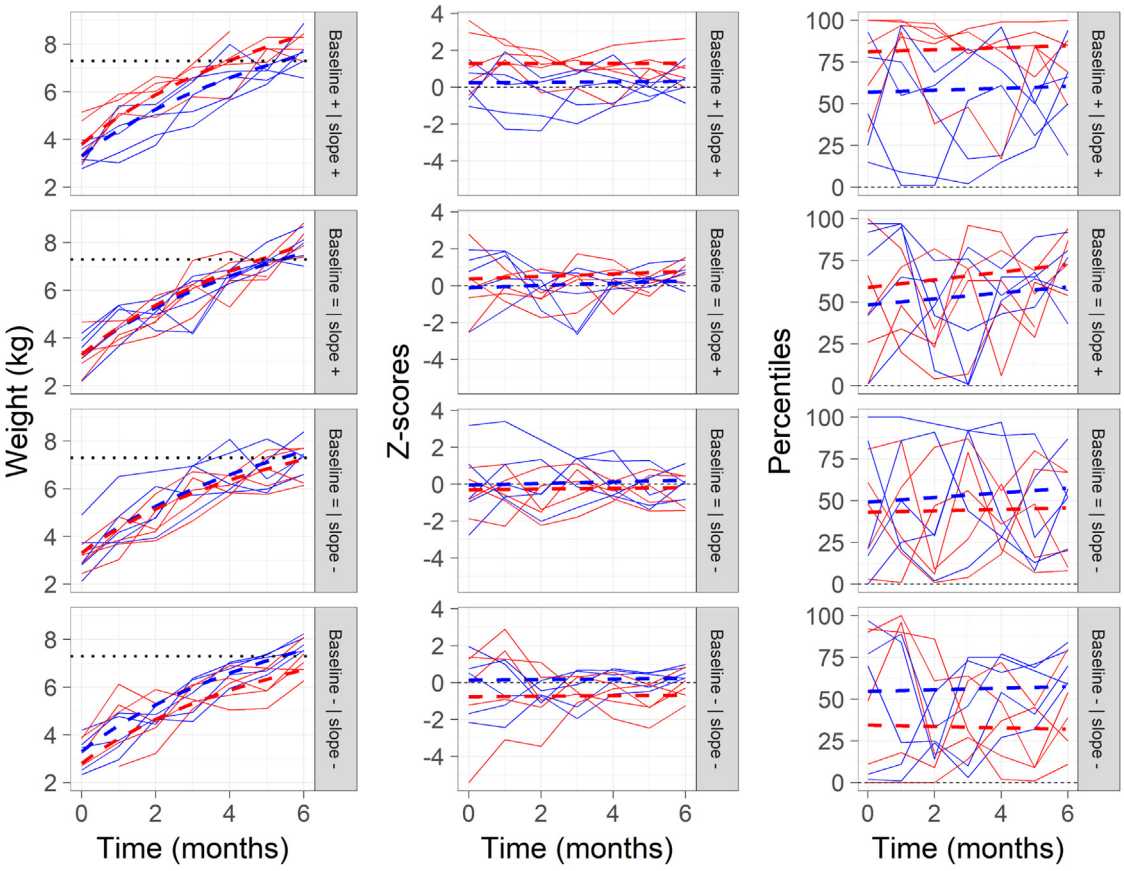

Fig. 1. Randomly selected simulated growth curves for Chilean females. Bold dashed lines represent the growth curve corresponding to simulation parameters: weight = 16.87 + 3.20 × t* − 8.96ln(t*) − 16.77/t* + β_4 × group + β_5 × group × t*, t* = (t + 9)/9, and group = 0 for reference and group = 1 for index exposure. Baseline + | Slope +: $\beta_4 = 0$, $\beta_5 = 0.5$; Baseline = | Slope +: $\beta_4 = -0.5$, $\beta_5 = 0.5$; Baseline = | Slope −: $\beta_4 = +0.5$, $\beta_5 = -0.5$; Baseline − | Slope −: $\beta_4 = 0$, $\beta_5 = -0.5$. Solid lighter lines represent the randomly selected simulated lines.

groups. Model 1 had time and intercept as random effects in analysis estimates:

$$y_{ij} = \beta_{0i} + \beta_{1i} \cdot month6 + \beta_2 \cdot group + \beta_3 \cdot month6 \cdot group + e_{ij}$$

An alternate model, model 2, had the same parameters as model 1, but with fixed effects and autocorrelated residuals.

Power estimates were from 1000 iterations in which the ANOVA model was fit and a binary indicator assigned to the *p-value* for the product term, β_3, indicating a significant difference in weight between the two groups for a one-unit change in time (6 months) at a designated α level of 0.05. Type I error was calculated from the same set of iterations under a null model. Finally, coefficients for the product term, β_3, were extracted from each iteration and a median and interquartile range of those values provided an estimated direction and variability of effect for each of the outcome measures.

We used R software for all analyses [8]. Simulations were conducted using R version 3.4.1 on a Linux-based computing system. Postsimulation data handling was done using RStudio version 1.0.136.

Results

We present the results for females with binary exposure proportion of 0.2 because interpretation of results remained the same across gender groups and exposure proportions. Estimates for males and those for exposure proportions of 0.5 are in the accompanying Appendix Tables 1–12. Similarly, the display of growth trajectories with different outcomes as presented in Figure 1 includes results for the Chilean sample as the comparisons were similar across the three different locations.

Visual inspection of the three outcome values in Figure 1 plotted against time for the unexposed group (matching the original parameter estimates in the literature) indicated distinct shapes dependent on the outcome. As specified for the simulated data, weight (in the first column of figures) increased over time with the exposed group clearly above and below the referent, in the first and last two panels, respectively. Weight converted to WHO Z-scores appeared to be flat and close to zero for the referent group, and the exposed group was also linear but above or below the referent by about 1 SD depending on the baseline weight status. Percentiles of the WHO Z-scores plotted over time indicated a referent group close to the 50th percentile, as expected, given the referent WHO Z-scores being close to 0. Similar to the WHO Z-scores, the exposed group percentile values were above and below the referent depending on the parameter combination in the data generation process.

Power

The use of untransformed weight values as the outcome provided the largest power estimates of the three outcome measures for all models considered (Table 1), exceeding 0.9 in all cases. In the circumstance of unequal baseline weight, the power was approximately double that of the transformed values. These results

Table 1
Estimated power by model*, country, and parameter combinations for females

Trajectory type	Sample	Weight		Z-score		Percentile	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Unequal baseline|+ slope	Portugal	0.991	1.000	0.230	0.472	0.266	0.042
	Italy	0.991	1.000	0.175	0.340	0.277	0.076
	Chile	0.989	1.000	0.272	0.455	0.223	0.021
Equal baseline|+ slope	Portugal	0.994	1.000	0.787	0.953	0.871	0.996
	Italy	0.995	1.000	0.805	0.951	0.884	0.998
	Chile	0.989	1.000	0.775	0.938	0.879	0.977
Equal baseline|− slope	Portugal	0.996	0.999	0.813	0.961	0.905	0.992
	Italy	0.994	1.000	0.841	0.963	0.881	0.990
	Chile	0.987	1.000	0.822	0.961	0.897	0.993
Unequal baseline|− slope	Portugal	0.989	1.000	0.146	0.418	0.324	0.581
	Italy	0.992	1.000	0.123	0.296	0.271	0.104
	Chile	0.990	1.000	0.162	0.391	0.348	0.222

Exposed proportion = 0.2 and $|\beta_5| = 0.5$.
* Model 1: random effects model and model 2: fixed effects model with autocorrelation residual structure.

spanned four different combinations of intercept/slope differences and two analytic models. For the simulation scenarios in which there were equal baseline weights across exposure groups, percentile outcome measures (range: 0.871–0.998) demonstrated favorable power estimates than most WHO Z-score values (range: 0.775–0.963). In contrast, the group of estimates with unequal baseline weights across exposure groups demonstrated that the estimated power for model 2 with WHO Z-score outcomes was higher than the corresponding model with percentile outcomes.

Type I error

As expected, most estimated alpha levels for simulated weight values were close to 0.05 (Table 2), the nominal value specified in generating the data. Fixed effects models (model 2) demonstrated lower type I errors for WHO Z-scores and percentiles compared to the original weight values, indicating a liberal bias. Otherwise, no clear patterns emerged when comparing the fitted models with a WHO Z-score or percentile outcome.

Product term estimates

In terms of product term estimates, the absolute value of the parameter for the time and group product term in the weight model, β_3 in models 1 or 2, was between 0.33 and 0.34 (Table 3). For simulated weight growth data having a positive or negative slope in the index group relative to the referent, the estimated product term was expected to be positive or negative, respectively, indicating increasing or decreasing growth of the index group relative to the referent group.

Given the different scales for each of the three outcomes, the estimated product term between time and group across different outcomes was not expected to have the same scale. However, the direction of the effect was expected to be the same. When examining the median product term estimates in certain models with WHO Z-score outcomes, it is clear these models consistently yielded estimates in the opposite direction than those for the observed weight outcome. Absent baseline weight differences, the product term coefficient estimates for the observed weight outcome models were similar to the WHO Z-scores outcome models. However, in every model with different baseline weight values in the index exposure group relative to the referent group, the median product term coefficient estimates were in the opposite direction of the true difference.

In our sensitivity analyses in which we altered the simulated weight growth differences to be even smaller than the ones described previously, we found similar differences conditional on outcome choice in type I error, power, and product term estimates across the exposure groups (Appendix).

Discussion

Our primary aim in this article was to determine the degree to which infant weight growth trajectory analyses are compromised by analytic limitations when using Z-scores or percentiles from a growth standard or reference in place of the observed value. To

Table 2
Estimated type I error estimates by model*, country, and parameter combinations for females

Trajectory type	Sample	Weight		Z-score		Percentile	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Unequal baseline|+ slope	Portugal	0.052	0.046	0.051	0.041	0.053	0.051
	Italy	0.055	0.046	0.047	0.037	0.039	0.042
	Chile	0.051	0.055	0.049	0.040	0.054	0.055
Equal baseline|+ slope	Portugal	0.053	0.042	0.054	0.027	0.054	0.034
	Italy	0.058	0.043	0.055	0.030	0.063	0.043
	Chile	0.051	0.051	0.041	0.041	0.042	0.042
Equal baseline|− slope	Portugal	0.041	0.041	0.050	0.034	0.056	0.040
	Italy	0.052	0.060	0.046	0.046	0.053	0.047
	Chile	0.049	0.037	0.040	0.034	0.048	0.036
Unequal baseline|− slope	Portugal	0.050	0.048	0.045	0.035	0.037	0.038
	Italy	0.055	0.049	0.052	0.037	0.047	0.037
	Chile	0.061	0.032	0.066	0.026	0.048	0.031

Exposed proportion = 0.2 and $|\beta_5| = 0.5$.
* Model 1: random effects model and model 2: fixed effects model with autocorrelation residual structure.

Table 3
Estimated time and exposure group product term median coefficient estimates, β_2 (interquartile range) by model*, country, and parameter combinations for females

Trajectory type	Sample	Weight		Z-score		Percentile	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Unequal baseline|+ slope	Portugal	0.33 (0.10)	0.33 (0.08)	−0.16 (0.18)	−0.20 (0.13)	2.73 (2.92)	0.75 (3.54)
	Italy	0.33 (0.10)	0.33 (0.08)	−0.15 (0.17)	−0.17 (0.13)	2.98 (2.65)	2.23 (3.43)
	Chile	0.34 (0.10)	0.34 (0.08)	−0.17 (0.19)	−0.19 (0.13)	2.44 (2.74)	−0.99 (3.08)
Equal baseline|+ slope	Portugal	0.34 (0.11)	0.33 (0.07)	0.39 (0.20)	0.37 (0.13)	6.82 (3.16)	12.87 (3.71)
	Italy	0.34 (0.10)	0.33 (0.08)	0.39 (0.20)	0.38 (0.14)	6.97 (2.84)	13.05 (3.79)
	Chile	0.34 (0.10)	0.33 (0.08)	0.36 (0.17)	0.35 (0.13)	6.65 (2.73)	11.61 (3.64)
Equal baseline|− slope	Portugal	−0.34 (0.10)	−0.33 (0.08)	−0.40 (0.19)	−0.39 (0.14)	−6.80 (2.89)	−12.81 (3.86)
	Italy	−0.33 (0.11)	−0.34 (0.07)	−0.40 (0.18)	−0.40 (0.13)	−6.78 (3.10)	−12.75 (3.67)
	Chile	−0.34 (0.10)	−0.33 (0.08)	−0.38 (0.18)	−0.37 (0.13)	−6.87 (2.81)	−12.87 (3.67)
Unequal baseline|− slope	Portugal	−0.33 (0.10)	−0.33 (0.08)	0.11 (0.19)	0.20 (0.16)	−3.13 (2.69)	−6.02 (3.03)
	Italy	−0.33 (0.10)	−0.33 (0.08)	0.10 (0.18)	0.16 (0.15)	−2.94 (2.63)	−2.63 (2.88)
	Chile	−0.33 (0.10)	−0.33 (0.08)	0.13 (0.20)	0.18 (0.14)	−3.29 (2.86)	−3.71 (3.37)

Exposed proportion = 0.2 and $|\beta_5| = 0.5$.
* Model 1: random effects model and model 2: fixed effects model with autocorrelation residual structure.

assess any analytic limitations, including power, type I error, and direction of effect, we conducted Monte Carlo simulations with infant growth data for three anthropometric measures: weight and the corresponding WHO Z-scores and percentiles. Within this framework, our interest focused on the growth rate differences between exposure groups. Given prior findings for an adolescent population, we expected lower power to detect differences in weight change by group with Z-scores compared to the original weight value.

Extensive Monte Carlo simulations confirmed the expected decrease in power when using Z-scores—previously demonstrated in an adolescent population [3]. In addition, there was the potential for conflicting inference under certain conditions when using WHO Z-scores for weight outcomes compared to the original non-transformed weight value. Direction switches in the estimated exposure group rate of weight change differences were unexpected findings across two outcome measures: weight Z-scores and observed weight value. Furthermore, these direction switches were consistent across sensitivity analyses with even smaller differences in birth weight and slopes across the exposure groups (Appendix).

One reason behind these unexpected findings could have been the choice of target population, the population to which inference is drawn [9,10]. A target population represented by the growth standard may have a different distribution of trajectories than the one represented by the source population. Analyzing those standardized values in a sequence of person-level longitudinal observations corresponds to change within that distribution of the standard, and inference with this new target population may no longer match changes in the study sample. Children with above-average birth weight in the standard may have a higher rate of weight change than the observed values in the source population, our simulated samples in this article. This difference between populations could lead to lower growth estimates than specified for the observed weight outcome after standardizing the weights. When we increased the specified rate of growth in the simulated source population, the resulting estimated median slope difference between exposure groups switched and was in the same direction, as hypothesized (data not shown). Considering that birth weight is correlated with slope of weight change, this finding would not be unusual.

These particular differences for standardized versus unstandardized outcomes may not be desirable as it relates to clinical or public health utility as different populations can have different growth parameters dependent on genetics and environment [11]. We assume that the target population in most research would be (1) nested in the source population from which the sample was drawn and (2) the potential subject for interventions. In our example, we used a WHO standard that reflects a standard for growth including breast-fed infants [12]. In turn, a mismatch between inference and the true target of intervention could lead to an unintended mismatch in policy recommendations. A consideration of the goals of the study may also be relevant to the choice of using observed weight versus standardized values of weight change. For example, if policy recommendations on growth will result from the study, the actual rate of growth in the sample, uncalibrated to a reference or standard, would be more informative.

In our simulated circumstances with exposure group comparisons of infants with different birth weights, the estimated differences in growth for the two exposure groups diverged for the original weight and WHO Z-scores. One reason supporting this finding could be the "true" growth distribution specified in the simulated study sample for the exposure group with above or below average birth weights did not match the growth distribution for those same children in the standard. In light of these findings, interpretation of results should emphasize the target population the researcher has chosen either via the standard or the original weight values. This interpretation strategy would account for any potential differences in estimates, given the choice of outcome measure and prevent recommendations based on Z-scores that may be irrelevant to the source population. If the target population is the source population from which the study sample was selected with an analytical goal to adjust for sex and age of the child, then a potential model in this context of longitudinal change would contain original weight outcomes stratified or conditioned on sex of the child.

Conclusions

Using Monte Carlo simulations, we have shown that using a Z-score standardized to an external population, such as a WHO standard, may have several inherent limitations in models of longitudinal weight change in infants. Simulations here provide evidence that analyses using standardized weight measures can (1) yield lower ability to detect real differences in the rate of growth between exposures groups, (2) identify groups with significant differences in which they do not exist, and (3) produce conclusions that may not match the true values for the source population. These three implications may impact research results to obscure true findings depending on the choice of target population. To improve interpretation of results from models of infant growth, we encourage reporting both standardized and original estimates

alongside careful consideration and clear identification of the target population when using standardized weight outcomes.

Acknowledgments

Ann Von Holle was supported by an American Heart Association predoctoral award (number 16PRE29200008).

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.annepidem.2018.04.006.

References

[1] Grummer-Strawn LM, Reinold C, Krebs NF. Centers for Disease Control and Prevention (CDC). Use of world health organization and CDC growth charts for children aged 0-59 months in the united states. MMWR Recomm Rep 2010;59(RR-9):1–15.

[2] Wang Y, Chen H-J. Use of percentiles and z-scores in anthropometry. In: Preedy VR, editor. Handbook of anthropometry. New York, NY: Springer New York; 2012. p. 29–48.

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

[4] Berkey CS, Reed RB. A model for describing normal and abnormal growth in early childhood. Hum Biol 1987;59(6):973–87.

[5] Hauspie RC, Cameron N, Molinari L. Methods in human growth research. Cambridge: Cambridge University Press; 2004.

[6] Pizzi C, Cole TJ, Corvalan C, Silva I dos S, Richiardi L, De Stavola BL. On modelling early life weight trajectories. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2014;177(2):371–96.

[7] WHO Multicentre Growth Reference Study Group, Onis M. Assessment of differences in linear growth among populations in the WHO multicentre growth reference study: Assessment of differences in linear growth. Acta Paediatrica 2007;95:56–65.

[8] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from https://www.r-project.org/.

[9] Rothman KJ, Greenland S. Modern epidemiology. Philadelphia, PA.: Wolters Kluwer/Lippincott Williams & Wilkins; 2008.

[10] Coggon D, Rose GA, Barker DJP. Epidemiology for the uninitiated. London: BMJ Books; 2003.

[11] Milani S, Buckler JMH, Kelnar CJH, Benso L, Gilli G, Nicoletti I, et al. The use of local reference growth charts for clinical use or a universal standard: A balanced appraisal. J Endocrinological Invest 2012;35(2):224–6.

[12] Onis M, de Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J. The WHO multicentre growth reference study: Planning, study design, and methodology. Food Nutr Bull 2004;25(1):S15–26.