Check for updates

Received: 31 May 2023 | Revised: 24 October 2023 | Accepted: 25 October 2023

DOI: 10.1002/oby.23957

PERSPECTIVE

Obesity A Research Journal | THE OBESITY SOCIETY | WILEY

Limitations of BMI *z* scores for assessing weight change: A clinical tool versus individual risk

Shana Adise[1] | Kyung E. Rhee[2] | Jennifer Laurent[3] | Elizabeth A. Holzhausen[4] | Panteha Hayati Rezvan[5] | Tanya L. Alderete[4] | Alaina P. Vidmar[1]

[1]Department of Pediatrics, Division of Endocrinology, Diabetes and Metabolism, Children's Hospital Los Angeles and Keck School of Medicine of USC, Los Angeles, California, USA

[2]Department of Pediatrics, University of California, San Diego, La Jolla, California, USA

[3]Department of Nursing, University of Vermont, Burlington, Vermont, USA

[4]Department of Integrative Physiology, University of Colorado Boulder, Boulder, Colorado, USA

[5]Children's Hospital Los Angeles, The Saban Research Institute, Biostatistics and Data Management Core, Los Angeles, California, USA

Correspondence
Shana Adise, Department of Pediatrics, Division of Endocrinology, Diabetes and Metabolism, Children's Hospital Los Angeles and Keck School of Medicine of USC, 4650 Sunset Blvd, Mailstop #61, Los Angeles, CA 90027, USA.
Email: sadise@chla.usc.edu

Funding information
National Heart, Lung, and Blood Institute, Grant/Award Number: T32 HL149646; National Institute of Diabetes and Digestive and Kidney Diseases, Grant/Award Numbers: K01 DK13584701, K23DK134801; National Institute of Environmental Health Sciences, Grant/Award Number: R01ES035035; National Institutes of Health and National Institute on Drug Abuse, Grant/Award Numbers: U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147; National Institute on Minority Health and Health Disparities, Grant/Award Number: P50MD017344; Sacchi Foundation, American Diabetes Assocation, Grant/Award Number: 11-22-ICTSN-32

Abstract

Although pediatric growth curves provide clinical utility, using these metrics for within-person change over time can be misleading. As research is focused on understanding cardiometabolic consequences of weight gain, it is important to use precise metrics to analyze these longitudinal research questions. Despite several foundational recommendations to limit the use of reference pediatric growth curves (e.g., BMI *z* scores) for within-person longitudinal research, it has evolved into the "gold standard" for using growth curves for pediatric weight gain analyses. Therefore, the objective of this paper is to discuss (A) the methodology used to create reference growth curves; (B) the appropriate use of reference pediatric BMI growth curves within the context of cross-sectional and longitudinal analyses in research; and (C) how to select metrics based on desired evaluations. Careful consideration using standardized references scores is essential when assessing obesity-related questions and comorbid risk over time in pediatric populations.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Authors. *Obesity* published by Wiley Periodicals LLC on behalf of The Obesity Society.

INTRODUCTION

Body mass index (BMI; kilograms/meters squared) is used to estimate fat mass when gold standard methods are not available [1, 2]. Calculated from BMI, weight status cutoffs (e.g., healthy weight, overweight, obesity) provide easy clinical screening for cardiometabolic risk and disease [3]. Unlike adults, specific BMI growth curves have been created for children to account for age- and sex-specific differences in growth [4, 5], where the BMI metric is often expressed as a *z* score. *Z* scores allow for insight into how a child is growing in reference to a population at a particular time, but there are statistical limitations for using within-person differences in *z* scores (e.g., z_{diff}) to assess weight gain accelerations that can extrapolate to miscalculations of cardiometabolic risk overtime. Across fields, using z_{diff} for longitudinal research has been highly cautioned against [6–10]. Yet, in pediatric weight gain research, this has been widely ignored. For example, a PubMed search for BMI *z* score change among infants, children, and adolescents (e.g., search term: "BMI z-score" AND change NOT cross-sectional [Title/Abstract]) from 2000 until present was associated with more than 1000 published articles. Therefore, the purpose of this paper is to (A) acknowledge the merit of pediatric BMI *z* scores for clinical practice and cross-sectional research and (B) discuss the limitations of z_{diff} for longitudinal research focused on understanding pediatric cardiometabolic risk. Careful selection of weight gain metrics is critical so that researchers and clinicians may translate research findings more appropriately.

REFERENCE SCORE CREATIONS AND CHALLENGES WITH USE FOR LONGITUDINAL ASSESSMENT

Reference, or *z*, scores are useful for classifying metrics of a population into categories to determine clinical pathology and outstanding or suboptimal performance. For example, they are useful for classifying limits for various domains, including scholastic achievement, intelligence, neurocognitive testing, and pediatric assessments of growth [6]. Reference scores are created using a transformation process called standardization, which transforms a raw score (e.g., BMI) into a relative score (e.g., BMI *z* score) and ranks individual scores according to others in a reference population (0–100, e.g., BMI percentile). Then, the reference population mean (m) is subtracted from each individual raw score (x) and divided by the reference population standard deviation (SD; z score $= (x - m)/SD$). For a normal distribution, a *z* score represents the number of SD a given value is from the population mean ($m = 0$) in the reference population (e.g., z score $= 1$ is equivalent to $+1$ SD above the population mean). Thus, a BMI *z* score is a measure of how much a child's raw BMI score deviates from the mean BMI for their age and sex compared with a standard reference population of other children of that age and sex

BMI *z* scores provide useful cross-sectional weight status classification cutoffs (e.g., healthy weight, obesity). However, this metric is not appropriate to assess within-person change over time because of its inherent dependency on a static underlying reference population [6, 10]. To better understand this concept, we provide a simple example of test scores of a class at Time 1 and Time 2, which occur before and after students learn about content relevant to the test. At Time 1, *Student A* obtained a raw score of 83 out of 100, where the class average was 80 (SD = 20). Because the teacher was interested in understanding how well the students performed and the test's difficulty (i.e., the distribution of the scores in reference to the class average), they standardized the scores (*Student A's* score fell into the 56th percentile, z score = 0.15). During the next assessment (i.e., Time 2), the class average dropped to 60 (SD = 13) but *Student A* received an 87 (98th percentile, z score = 2.08). Next, the teacher was interested in understanding if students improved on performance from their first to second test (i.e., within-person change, pre- vs. post-test assessment). If the teacher used the z_{diff} to assess within-person change, this could provide a skewed perspective. For example, for *Student A*, their z_{diff} would be 1.90, which suggests that they had a 42% improvement in comparison to the change among the entire class. Meanwhile, in reference to themself, *Student A* only had a modest 4-point improvement between the tests, suggesting that for this student, the instructional unit at hand did not substantially improve their score by 42%. Thus, if the goal was to track *Student A*'s performance over time, it would be more appropriate to use their actual raw scores.

The example highlights how z_{diff} can lead to misleading longitudinal results, but in the case of BMI *z* score, this is even more problematic as the standardized variable is time-locked to a specific stagnant population. For example, BMI *z* scores for 9-year-olds were sampled from a different population than the 12-year-olds, distribution. Therefore, examining *z* score changes would provide information with respect to underlying reference populations at 9 and 12 years, and not within a single child. Thus, this creates an extreme bias for longitudinal research focused on examining within-child weight gain. Figure 1 provides a visualization of this dilemma using data from youth enrolled in The Adolescent Brain Cognitive Development (ABCD) Study (www.ABCDStudy.org) [11, 12]. Here, we were interested in examining the relationship between brain development and extreme weight gain among youth with a stable weight compared with those without weight stability (i.e., extreme weight gain) from 9/10 to 11/12 years old. Initially, using clinical practice guidelines, weight stability was defined as $z_{diff} < 0.2$ and nonstability as $z_{diff} > 0.2$ [13]. This criterion grossly overestimated non-weight stability as 52% of youth were classified with nonstable weight, of whom 79.3% had a healthy weight at baseline (Figure 1A). Notably, 74.3% of these youth were still classified as having a healthy weight 2 years later. Thus, we compared how well this criterion mapped onto physical weight gain, and not surprisingly, this metric performed poorly again (Figure 1B). There was no specificity to identify youth who had weight gain beyond $+1$ SD above the mean (dashed black line). As one solution, weight stability was expanded to only include those of a healthy weight, whereas weight nonstability included gaining >38 lb in 2 years ($+1$ SD above the mean; Figure 1C). Importantly, this grouping was not confounded by height, suggesting that those identified as having extreme weight gain did not also have a growth

1930739x, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.23957, Wiley Online Library on [28/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Obesity A Research Journal THE OBESITY SOCIETY WILEY

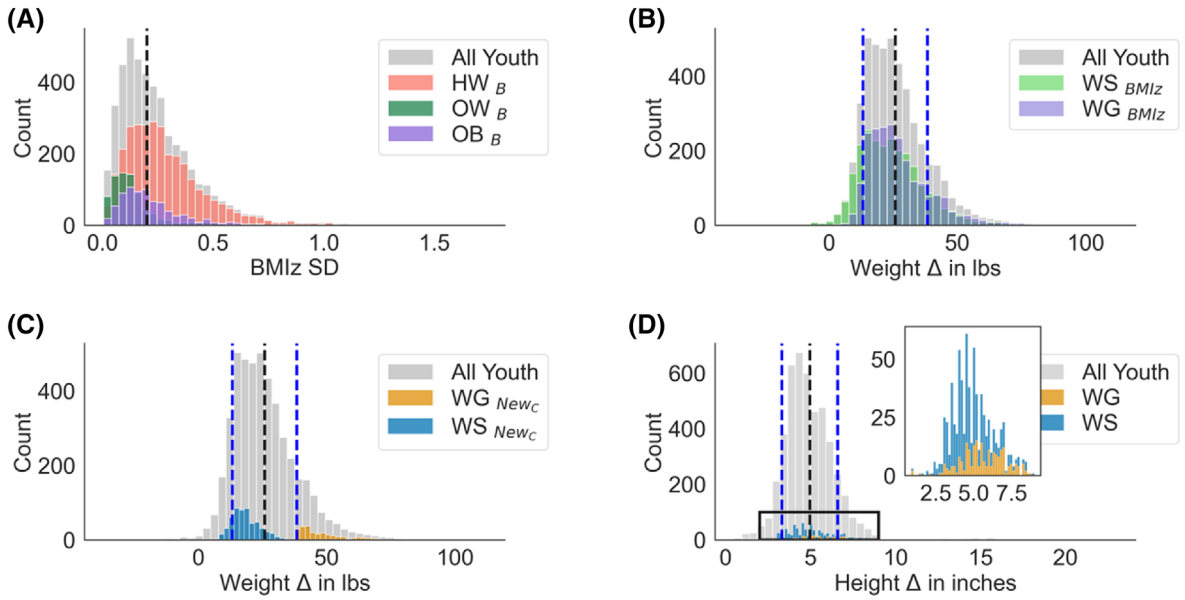


FIGURE 1 (A) Changes in BMI *z* score SD across two time points (ages 9/10 years old to 11/12 years old, within-person). The dashed lined indicates the clinical cutoff for weight stability using BMI *z* score SD changes, where <0.2 indicates weight stability. This figure depicts that most children were not classified as weight stable. HW, healthy weight; OW, overweight; OB, obesity, at baseline (B; ages 9/10 years old). (B) Change in physical weight between the two time points for each group using the clinical criterion of weight stability (e.g., weight stable [WS], weight gain [WG]). Dashed black line corresponds to the average weight gain across the sample, where the blue dashed lines depict −1 and +1 SD above the mean. (C) Distributions of weight gain by the new classification that included a weight gain criterion (>38 lb of weight gain across 2 years, plus the clinical BMI *z* score cutoff; New_C). (D) Demonstration that weight stable and weight gain groups did not differ in height velocity. Reproduced with permission from Adise et al. [12]. [Color figure can be viewed at wileyonlinelibrary.com]

spurt (Figure 1D). This classification definition seemed to better capture extreme weight gain, which was associated with underlying differences in brain structure prior to weight gain and measurable changes in brain structure after weight gain. Although imperfect, this example highlights how individual weight change may be more relevant to biological outcomes when compared with changes in reference growth charts.

ALTERNATIVE APPROACHES FOR LONGITUDINAL ANALYSES

Although there are better metrics to assess adiposity over time (e.g., DXA, MRI, waist circumference, bioelectrical impedance), BMI remains a quick and cost-effective tool for many researchers and physicians alike. Therefore, it is unrealistic to completely forgo the use of BMI *z* scores, and the most accurate metric for assessing weight change in children is debated [7, 11, 12]. Some guidance exists for pediatric clinical trial weight reduction outcomes (e.g., mean change in BMI, proportion of patients who lose greater than or equal to 5% of baseline), but there is no common consensus. There are valid concerns regarding the use of raw BMI for pediatric longitudinal analyses, but statistical models can account for age- and sex-growth patterns, like modeling continuous norming [14, 15]. In adults, growing evidence suggests that a 5% change in actual weight may be a good surrogate marker to capture longitudinal changes in adiposity and cardiometabolic risk [16, 17]. However, there is limited research conducted in children to support this. Perhaps an alternative approach would be to examine weight change while adjusting for sex, height, and age, rather than using BMI *z* scores. Alternatively, bioelectrical impedance is an alternative time and cost-effective tool to measure change in adiposity in youth; however, there is no consensus on how the output metrics obtained from these tests predict risk overtime and thus their practical clinical relevance is often questioned. Currently, it is challenging to recommend one longitudinal pediatric weight-based metric over another, leaving the researcher to decide the best measure for the study and outcome of interest. However, we contend that BMI *z* scores are best for cross-sectional classification purposes, and it should be highly cautioned against to use z_{diff} for longitudinal outcomes.

ADDITIONAL LIMITATIONS OF GROWTH CURVES

It is also important to highlight the inherent racial biases specifically in the Centers for Disease Control and Prevention (CDC) *z* scores, which were originally created using data pooled from the National Health and Nutritional Examination Survey (NHANES) from 1963 to 1980, a largely White and educated sample. In 2022, the CDC updated the reference tables with data collected in the last decade [18]. However, data from 1963 to 1980 were still used for estimation, perpetuating racial and socioeconomic biases [19], as children from

1930739x, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.23957, Wiley Online Library on [28/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

diverse backgrounds are at risk for inappropriately weight classification [20-22]. One solution may be to redefine the CDC growth charts with more recent data. Yet, this recommendation should be proceeded with caution as obesity rates have drastically increased [23, 24], and if *z* scores are used to define acceptable weight parameters (e.g., healthy vs. overweight), youth currently classified as having obesity may then be erroneously categorized as having a healthy weight. Although other reference charts (e.g., World Health Organization's international score) [4] may circumvent some of these sociodemographic issues, the misuse of the *z* score statistical tool in longitudinal scientific questions is still notable. Moreover, there are wide variations in body weight distribution by race and ethnicity [25], which means these cutoff points may not accurately capture metabolic risk accordingly. As such, there is a dire need for science to redefine weight-related cut-points for assessing excess adiposity with respect to encompassing individualistic characteristics (e.g., height, weight, age, sex, race/ethnicity) rather than relying on outdated and biased population estimates that provide little insight into a child's metabolic health.

CALL TO SCIENTIFIC COMMUNITY

There is an urgent need for further investigation into precise, easy, and cost-effective metrics that can accurately assess weight-related metabolic risk. Pediatric clinicians spend on average 15 min with a patient, necessitating quick, precise estimates of weight gain velocities that are useful in evaluating and managing care. Research settings are often afforded the time to conduct nuanced and precise analyses and thus can inform efficient metrics for estimating weight change that are clinically useful and metabolically meaningful. Perhaps with deliberate and intentional use of algorithms with less bias, research will be better translated to the clinical settings, resulting in more accurate measurements of within-person weight gain velocities that are more indicative of cardiovascular risk than inaccurate z_{diff} estimates. We would like to also emphasize the importance of collaboration between statisticians and subject-matter experts for successful data analysis. Such engagement and collaboration would lead to more robust conclusions and ultimately contribute to better decision-making in the clinical setting. Taken together, we ask the scientific and clinical communities to do the following: (A) acknowledge the limitations in using z_{diff} (and percentile differences) for longitudinal analyses, (B) update interpretation of anthropometric metrics in clinical practice, (C) identify software that can be more systematically incorporated into electronic health records to provide clinicians will real-time, accurate assessment of anthropometric measures in youth, and (D) understand that the current BMI *z* scores are subject to bias, especially among historically excluded groups with the greatest increase in prevalence of pediatric obesity.O

AUTHOR CONTRIBUTIONS

Shana Adise and Alaina P. Vidmar conceptualized the manuscript. Shana Adise, Kyung E. Rhee, Jennifer Laurent, Elizabeth A. Holzhausen, Tanya L. Alderete, Panteha Hayati Rezvan, and Alaina P. Vidmar drafted the initial manuscript and reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Michael I. Goran for his support throughout various stages of this manuscript preparation.

FUNDING INFORMATION

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development[SM] Study (https://abcdstudy.org/), held in the NIMH Data Archive (NDA). The ABCD Study is supported by the National Institutes of Health (NIH) and National Institute on Drug Abuse and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U01DA050989, U01DA051016, U01DA051018, U01DA051037, U01DA051039, U01DA051038, U01DA050987, U01DA050988, U01DA041025, U01DA041093, and U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners/. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/principal-investigators/. The ABCD Study consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or other ABCD Study consortium investigators. The ABCD Study data repository grows and changes over time. The ABCD Study data used in this report came from https://doi.org/10.15154/1503209. Shana Adise and Alaina P. Vidmar were supported by the National Institute of Diabetes and Digestive and Kidney Diseases award numbers K01 DK13584701 and K23DK134801. Alaina P. Vidmar was also supported by the Sacchi Foundation and American Diabetes Association award number 11-22-ICTSN-32. Elizabeth A. Holzhausen was supported by the National Heart, Lung, and Blood Institute award number T32 HL149646. Tanya L. Alderete was supported by the National Institute on Minority Health and Health Disparities award number P50MD017344 and from the National Institute of Environmental Health Sciences award number R01ES035035.

CONFLICT OF INTEREST STATEMENT

The authors declared no conflict of interest.

ORCID

Shana Adise https://orcid.org/0000-0002-7225-7724

REFERENCES

1. Gradmark AMI, Rydh A, Renström F, et al. Computed tomography-based validation of abdominal adiposity measurements from ultrasonography, dual-energy X-ray absorptiometry and anthropometry. *Br J Nutr*. 2010;104(4):582-588.

19307939x, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.23957, Wiley Online Library on [28/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2. Martin-Calvo N, Moreno-Galarraga L, Martinez-Gonzalez MA. Association between body mass index, waist-to-height ratio and adiposity in children: A systematic review and meta-analysis. *Nutrients*. 2016;8(8).
3. Berker D, Koparal S, Işik S, et al. Compatibility of different methods for the measurement of visceral fat in different body mass index strata. *Diagn Interv Radiol*. 2010;16(2):99-105.
4. World Health Organization. Child growth standards. https://www.who.int/tools/child-growth-standards
5. Kuczmarski RJ, Ogden CL, Guo SS, et al. 2000 CDC growth charts for the United States: Methods and development. Vital and Health Statistics, series 11, no. 246. National Center for Health Statistics; 2002.
6. Moeller J. A word on standardization in longitudinal studies: Don't. *Front Psychol*. 2015;6:1389. doi:10.3389/fpsyg.2015.01389
7. Cole TJ, Faith MS, Pietrobelli A, Heo M. What is the best measure of adiposity change in growing children: BMI, BMI %, BMI z-score or BMI centile? *Eur J Clin Nutr*. 2005;59(3):419-425.
8. 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.
9. Hendrickson MA, Pitt MB. Three areas where our growth chart conversations fall short-Room to grow. *JAMA Pediatr*. 2022;176:123-124. doi:10.1001/jamapediatrics.2021.4330
10. Maloney ES, Larrivee LS. Limitations of age-equivalent scores in reporting the results of norm-referenced tests. *Contemp Issues Commun Sci Disord*. 2007;34:86-93.
11. Adise S, Allgaier N, Laurent J, et al. Multimodal brain predictors of current weight and weight gain in children enrolled in the ABCD Study®. *Dev Cogn Neurosci*. 2021;49:100948. doi:10.1016/j.dcn.2021.100948
12. Adise S, Marshall AT, Hahn S, et al. Longitudinal assessment of brain structure and behaviour in youth with rapid weight gain: Potential contributing causes and consequences. *Pediatr Obes*. 2023;18:e12985. doi:10.15154/1503209
13. Geserick M, Vogel M, Gausche R, et al. Acceleration of BMI in early childhood and risk of sustained obesity. *N Engl J Med*. 2018;379(14):1303-1312.
14. Hosey CM, Halpin K, Shakhnovich V, et al. Pediatric growth patterns in youth-onset type 2 diabetes mellitus: Implications for physiologically-based pharmacokinetic models. *Clin Transl Sci*. 2022;15(4):912-922.
15. Voncken L, Albers CJ, Timmerman ME. Bias-variance trade-off in continuous test norming. *Assessment*. 2021;28(8):1932-1948.
16. Ryan DH, Yockey SR. Weight Loss and improvement in comorbidity: differences at 5%, 10%, 15%, and over. *Curr Obes Rep*. 2017;6:187-194.
17. Hampl SE, Hassink SG, Skinner AC, et al. Clinical practice guideline for the evaluation and treatment of children and adolescents with obesity. *Pediatrics*. 2023;151(2).
18. Hales CM, Freedman DS, Akinbami L, Wei R, Ogden CL. Evaluation of alternative body mass index (BMI) metrics to monitor weight status in children and adolescents with extremely high BMI using CDC BMI-for-age growth charts. Vital and Health Statistics, series 2, no. 197. National Center for Health Statistics; 2022.
19. Ryder JR, Kelly AS, Freedman DS. Metrics matter: Toward consensus reporting of BMI and weight-related outcomes in pediatric obesity clinical trials. *Obesity*. 2022;30(3):571-572.
20. Jih J, Mukherjea A, Vittinghoff E, et al. Using appropriate body mass index cut points for overweight and obesity among Asian Americans. *Prev Med*. 2014;65(1):1-6.
21. Kam LY, Huang DQ, Teng MLP, et al. Clinical profiles of Asians with NAFLD: A systematic review and meta-analysis. *Dig Dis*. 2022;40(6):734-744.
22. Koirala B, Turkson-Ocran RA, Baptiste D, et al. Heterogeneity of cardiovascular disease risk factors among Asian immigrants: Insights from the 2010 to 2018 national health interview survey. *J Am Heart Assoc*. 2021;10(13).
23. Fryar CD, Carroll MD, Ogden CL. Prevalence of overweight, obesity, and severe obesity among adults aged 20 and over: United States, 1960-1962 through 2015-2016. Health E-Stats. National Center for Health Statistics; 2018.
24. Ogden CL, Fryar CD, Hales CM, Carroll MD, Aoki Y, Freedman DS. Differences in obesity prevalence by demographics and urbanization in US children and adolescents, 2013-2016. *JAMA*. 2018;319(23):2410-2418.
25. Liu B, Du Y, Wu Y, Snetselaar LG, Wallace RB, Bao W. Trends in obesity and adiposity measures by race or ethnicity among adults in the United States 2011-18: Population based study. *BMJ*. 2021;372:n365. doi:10.1136/bmj.n365

How to cite this article: Adise S, Rhee KE, Laurent J, et al. Limitations of BMI *z* scores for assessing weight change: A clinical tool versus individual risk. *Obesity (Silver Spring)*. 2024;32(3):445-449. doi:10.1002/oby.23957

1930739x, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.23957, Wiley Online Library on [28/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License