As the prevalence of pediatric severe obesity continues to increase, it has become apparent that some of the “tried and true” body mass index (BMI) metrics many researchers and clinicians have become accustomed to using, such as BMI z-score, are proving less reliable in certain circumstances and in many cases can be misleading. Two reports from Freedman et al, the most recent published in the current volume of The Journal, raise important concerns about the use of BMI z-score in studies that include children with severe obesity. Using data from the United States Centers for Disease Control and Prevention, Pediatric Nutrition Surveillance System, the authors analyzed BMI values from 8.7 million children ages 2-4 years old obtained between 2008 and 2011. They demonstrated that the theoretical maximum BMI z-score varied by more than 3-fold across ages and in children with severe obesity, BMI z-score was only moderately correlated with BMI percent of the 95th percentile (ie, the percentage above or below the absolute BMI cutpoint associated with the age- and sex-specific 95th BMI percentile) and delta BMI percent of the 95th percentile (ie, the difference in absolute BMI units of the child’s BMI from the 95th percentile BMI). They also demonstrated that BMI z-score values could differ by more than 1 SD among children with similar body size, simply owing to differences in age or sex. These findings extend recent observations from this same group showing almost no association between BMI z-score and measures of body fatness in children and adolescents with severe obesity. Together, these studies draw the same conclusion using 2 distinct datasets, namely, that BMI z-score (generated from the Centers for Disease Control and Prevention growth charts) is a poor metric when applied to youth with severe obesity because values do not correlate with adiposity level and can be highly misleading, potentially causing erroneous conclusions to be drawn from research studies and even in the clinical setting.

Freedman et al cite a specific example of how a weight management intervention implemented among 2-year-old girls with severe obesity might seem to be effective based on change in BMI z-score when in fact the intervention may have had no impact on the true BMI trajectory, simply owing to mathematical limitations regarding how BMI z-score is calculated among girls in this age group. The authors contend that the exact opposite conclusion could be drawn if the same intervention were to be evaluated in 2-year-old boys. In these types of scenarios, BMI z-score is applied essentially in a way in which it was never intended to be used and, accordingly, generates spurious results. As an aside, it should be noted that BMI z-scores derived from international cohorts, such as from Britain and The Netherlands, may be less susceptible to the problems identified with BMI z-scores using the Centers for Disease Control and Prevention growth chart data, and therefore could potentially have utility in the context of severe obesity.

Freedman et al and others have suggested that alternative metrics, such as BMI percent of the 95th percentile and delta BMI percent of the 95th percentile, are much more flexible and robust in the context of severe obesity as compared with BMI z-score. Although these measures seem to be more appropriate from an arithmetic perspective and likely preferred to BMI z-score, further research is needed to evaluate their usefulness by characterizing the association of changes in these measures with changes in adiposity in the context of longitudinal studies and interventional trials, as well as in identifying what magnitude of reduction is necessary to achieve clinically meaningful improvements in obesity-related cardiometabolic risk factors and comorbidities. For example, in the adult realm, it is now widely accepted that 3%-5% body weight reduction is clinically meaningful, whereas an equivalent threshold of weight loss (BMI reduction) has yet to be identified for children or adolescents, perhaps because so many different BMI metrics have been used with no consensus on which is the preferred measure(s).

So, what are researchers and clinicians to make of this and how should they respond? Although BMI z-score remains useful for obesity classification and BMI tracking among youth ranging from normal weight to moderate obesity, a convincing case has now been made to move away from the use of BMI z-score as a primary weight loss outcome in research involving youth with severe obesity and as a metric of tracking change among individual patients with high BMI values in the clinical setting. Researchers should carefully select which BMI measures to incorporate into study protocols based on the degree of obesity of the target population and health systems should implement alternative ways to track BMI trajectories within the electronic medical record that more accurately reflect body size fluctuations in youth with high BMI. In this time of relative uncertainty as to which BMI measures are the most accurate and useful, researchers should err on the side of reporting more, not fewer, BMI-derived outcomes in their publications to offer readers more flexibility in how to interpret findings from individual studies and allow for better comparison of results across regions.
studies. Peer reviewers and journal editors should push back on authors who report BMI z-score as a primary outcome in studies involving participants with severe obesity and insist that alternative BMI metrics be presented in publications.

In response to the increasing number of children and adolescents with severe obesity, a logical next step for the scientific and clinical community is to work toward consensus on preferred BMI-based measures to be used uniformly in research studies involving participants with high BMI values and for systematic implementation in the clinical setting for classification and tracking purposes. Flexible metrics that can be applied to the entire range of the BMI spectrum are desirable so that direct comparisons can be drawn across different studies involving individuals with widely varying BMI levels. Currently, the field is hamstrung by a “Tower of Babel” phenomenon in which too many different “BMI languages” are being spoken, some of which (eg, BMI z-score) should no longer be uttered in certain settings. The field needs to establish a common dialect, shared by the research and clinical community alike, in an effort to maximize efficiencies and accelerate progress for one of the most pressing medical and public health issues of our time.

Aaron S. Kelly, PhD
Departments of Pediatrics and Medicine
University of Minnesota Medical School, and
University of Minnesota Masonic Children’s Hospital
Minneapolis, Minnesota

Stephen R. Daniels, MD, PhD
Department of Pediatrics
University of Colorado School of Medicine, and
Children’s Hospital Colorado
Aurora, Colorado

References