



Published in final edited form as:

Child Obes. 2025 January ; 21(1): 22–29. doi:10.1089/chi.2024.0248.

High BMI z-Scores from Different Growth References Are Not Comparable: An Example from a Weight Management Trial with an Anti-Obesity Medication in Pubertal Adolescents with Obesity

Craig M. Hales, MD, MPH, MS^{1,*}, Cynthia L. Ogden, PhD¹, David S. Freedman, PhD², Kushal Sahu, M.Sc³, Paula M. Hale, MD⁴, Rashmi K. Mamadi, MD³, Aaron S. Kelly, PhD⁵

¹Centers for Disease Control and Prevention, Hyattsville, MD, USA.

²Centers for Disease Control and Prevention, Atlanta, GA, USA.

³Novo Nordisk Global Business Services, Bengaluru, Karnataka, India.

⁴Novo Nordisk Inc., Plainsboro, NJ, USA.

⁵Department of Pediatrics and Center for Pediatric Obesity Medicine, University of Minnesota Medical School, Minneapolis, MN, USA.

Abstract

Background: The BMI z-score is a standardized measure of weight status and weight change in children and adolescents. BMI z-scores from various growth references are often considered comparable, and differences among them are underappreciated.

Methods: This study reanalyzed data from a weight management clinical study of liraglutide in pubertal adolescents with obesity using growth references from CDC 2000, CDC Extended, World Health Organization (WHO), and International Obesity Task Force.

Results: BMI z-score treatment differences varied 2-fold from −0.13 (CDC 2000) to −0.26 (WHO) overall and varied almost 4-fold from −0.05 (CDC 2000) to −0.19 (WHO) among adolescents with high baseline BMI z-score.

Conclusions: Depending upon the growth reference used, BMI z-score endpoints can produce highly variable treatment estimates and alter interpretations of clinical meaningfulness. BMI z-scores cited without the associated growth reference cannot be accurately interpreted.

Address correspondence to: Craig M. Hales, MD, MPH, MS, Centers for Disease Control and Prevention, 3311 Toledo Rd, Hyattsville, MD 20892, USA, craig.hales@nih.gov.

*Current Address: National Institutes of Health, Bethesda, MD, USA.

Authors' Contributions

C.M.H.: Conceptualization; writing—original draft; writing—reviewing and editing; visualization. C.L.O.: Supervision; writing—reviewing and editing. D.F.: Writing—reviewing and editing. K.S.: Data curation; formal analysis. P.M.H.: supervision; writing—reviewing and editing. R.K.M.: Data curation; formal analysis. A.S.K.: Conceptualization; writing—reviewing and editing.

Author Disclosure Statement

C.M.H., C.L.O., and D.F. report no conflicts of interest. K.S., P.M.H., and R.K.M. are employees of and hold shares in Novo Nordisk. A.S.K. engages in unpaid consulting and educational activities for Boehringer Ingelheim, Eli Lilly, Novo Nordisk, and Vivus and receives donated drug/placebo from Vivus and donated drug from Novo Nordisk for National Institute of Diabetes and Digestive and Kidney Diseases-funded clinical trials.

Disclaimer

The findings and conclusions in this report are those of the authors and not necessarily represent the official position of the CDC.

Keywords

BMI; childhood obesity; clinical endpoints; clinical trials; growth chart; growth reference

Introduction

The change in BMI z-score is considered to be a standardized measure of weight change in children and adolescents.¹ Studies have attempted to quantify a threshold of BMI z-score change required for meaningful improvements in obesity-related biomarkers and health outcomes,^{2–5} although no consensus has yet emerged. The US FDA and the European Medicines Agency (EMA) recommend a primary endpoint based on BMI for clinical studies of anti-obesity medications in pediatric populations, and the EMA specifically recommends the change in BMI z-score as the primary endpoint.^{6,7}

Several national and international BMI-for-age growth references have been developed for applications in diverse populations and settings, resulting in a variety of reference populations and statistical methods being used for describing BMI distributions. BMI z-scores from different growth references are often considered comparable. For example, meta-analyses of obesity interventions^{4,5} often pool results from studies in which BMI z-scores are reported from a variety of different growth references without a critical evaluation of the implications. The US Preventive Services Task Force (USPSTF) relied, in part, on a meta-analysis of studies that used diverse growth references, such as CDC 2000, World Health Organization (WHO), and International Obesity Task Force (IOTF), for its recommendations for obesity screening and for comprehensive, intensive behavioral interventions in children and adolescents aged 6 years and older. The USPSTF concluded that a 0.15–0.25 reduction in BMI z-score is associated with improvements in cardiovascular and metabolic risk factors without citing any particular growth reference.¹

The comparability among growth references for children and adolescents with obesity is uncertain. The goal of this study was to evaluate the impact of using BMI z-scores from different growth references on clinical study results by reanalyzing data from a single weight management clinical study in pubertal adolescents with obesity. Study results were compared using CDC 2000, CDC Extended, WHO, and IOTF BMI-for-age growth references. Percentile and z-score reference curves, as well as plots showing the relationship between BMI and BMI z-score using these four growth references, were explored to help in explaining the observed variability.

Methods

This was a secondary analysis of a clinical study ([Clinical-Trials.gov](https://clinicaltrials.gov/ct2/show/study/NCT02918279) number [NCT02918279](https://clinicaltrials.gov/ct2/show/study/NCT02918279)) comparing liraglutide 3.0 mg once daily with placebo over 56 weeks among 251 adolescents aged 12–17 years with obesity.⁸ The primary endpoint of the original study was treatment difference in terms of BMI z-score (using the WHO growth reference) between liraglutide and placebo groups. For the current study, baseline weight status and change in BMI were expressed in terms of BMI (kg/m²) and BMI z-score with respect to four growth references: CDC 2000,⁹ CDC Extended,¹⁰ WHO,¹¹ and IOTF.¹² Although

the primary analysis did not evaluate efficacy stratified by baseline weight status, the current analysis used the overall mean baseline BMI z-score (WHO) of 3.17 to divide study participants into high (≥ 3.17) baseline BMI z-score and low (< 3.17) baseline BMI z-score subgroups to evaluate differences in the four growth references at high versus low BMI ranges. Differences between BMI change in liraglutide and placebo groups were calculated using mixed-effects models for repeated measurements according to intention-to-treat principles and included visit, treatment assignment, sex, region, baseline glycemic category, stratification factor for Tanner stage and interaction between baseline glycemic category and stratification factor for Tanner stage, baseline BMI z-score, and baseline age. All these factors and covariates were nested under visit, which is technically the same as introducing the corresponding interaction terms in the model. A compound symmetry covariance matrix for the BMI z-score measurements within the subject was used.

To better understand how the four different growth references lead to different BMI z-score changes, graphic growth charts were shown with reference curves at overweight and obesity cutoffs [i.e., 85th and 95th percentiles for CDC 2000 and CDC Extended, z-scores 1 (Z1) and 2 (Z2) for WHO, and IOTF-25 and IOTF-30 for IOTF] and Z2 through Z5 in accordance with the methods for each growth reference.^{9–12} Although the CDC 2000 reference is not recommended for use in children and adolescents with obesity,¹⁰ it is included here to illustrate the potential measurement bias in previously published meta-analyses that incorporated studies using the CDC 2000 reference. The four growth references span different age ranges: CDC 2000 and CDC Extended from 2 to 20 years, WHO from 5 to 19 years, and IOTF from 2 to 18 years. The relationships between BMI and BMI z-score using each of the four growth references were plotted for boys and girls at age 14.5 years (174 months), the mean baseline age of study participants.⁸ Treatment differences were considered statistically significant at $\alpha = 0.05$.

Analyses were performed using R¹³ and SAS Enterprise Guide (v8.3). BMI z-score (WHO) was calculated using the AnthroPlus R package.¹⁴

Results

The study included 251 adolescents (125 liraglutide and 126 placebo), of which 59.4% were female. The mean age was 14.5 years, with a mean BMI of 35.6 kg/m² and a mean BMI z-score (WHO) of 3.17 (> 99.9 th percentile) at baseline.⁸ Over the 56-week observation period, the average absolute BMI change was -1.4 kg/m² (-4.4%) in the liraglutide group and 0.4 kg/m² (0.9%) in the placebo group (Table 1), with a treatment difference of -1.8 kg/m². The treatment difference was -1.5 kg/m² in the high baseline BMI z-score group and -1.9 kg/m² in the low baseline BMI z-score group.

Figure 1 shows mean baseline BMI z-scores overall and for adolescents in high and low baseline BMI z-score groups. The mean baseline BMI z-score ranged from 2.3 (CDC 2000) to 3.2 (WHO) overall. The spread of baseline BMI z-scores across the four growth references was different between the high and low baseline BMI z-score groups. In the high baseline BMI z-score group, the BMI z-score value was lowest at 2.6 (CDC 2000) and was highest (WHO) at 3.7 (a spread of 1.1 kg/m²), whereas, in the low baseline BMI z-score

group, the BMI z-score was lowest at 2.0 (CDC Extended) and highest (WHO) at 2.7 (a spread of 0.7 kg/m²).

Change in BMI z-score from baseline to week 56 ranged from -0.23 (WHO and IOTF) to -0.14 (CDC Extended) in the liraglutide group and from -0.03 (CDC 2000 and IOTF) to 0.06 (CDC Extended) in the placebo group (Fig. 2). The four growth references showed different patterns of BMI z-score change from baseline to week 56 in the high baseline BMI z-score group compared with adolescents with low baseline BMI z-score and in the liraglutide group compared with the placebo group. For the placebo group, the BMI z-score change was negative for CDC 2000 and IOTF, and it was positive for CDC Extended and WHO overall and in the high baseline BMI z-score group.

Overall treatment difference in terms of BMI z-score between liraglutide and placebo groups (*i.e.*, the primary endpoint) varied from -0.26 (WHO) to -0.21 (CDC Extended) to -0.19 (IOTF) and to -0.13 (CDC 2000) (Fig. 2). Treatment difference for all four growth references were statistically significant ($p < 0.05$). The highest magnitude treatment differences occurred with the WHO BMI z-score in the high and low baseline BMI z-score groups, (-0.19 and -0.30, respectively), and the lowest magnitude occurred with the CDC 2000 BMI z-score (-0.05 and -0.19, respectively). All treatment differences in the lower baseline BMI z-score were statistically significant, whereas in the higher baseline BMI z-score group, only the CDC Extended treatment difference was statistically significant.

To illustrate the underlying differences in the four growth references, Figures 3A–D present four graphical BMI-for-age growth charts for boys and girls, with reference curves indicating cutoffs for overweight and obesity as well as z-scores 2 through 5 and a vertical reference line at age 14.5 years indicating the mean age of study participants. Subplots of the relationship between BMI and BMI z-score for each of the growth references at age 14.5 years are shown above each chart. Compression of z-scores at high values is observed for the CDC 2000 (Fig. 3A) and IOTF (Fig. 3D) growth references. Curves for Z4 in the WHO (Fig. 3C) and IOTF (Fig. 3D) references illustrate paradoxically decreasing values with increasing age in late adolescence.

Discussion

In this reanalysis of a clinical study evaluating the effectiveness of an anti-obesity medication among adolescents, variation in BMI z-score change in liraglutide and placebo groups resulted in BMI z-score treatment differences that varied 2-fold from -0.13 (CDC 2000) to -0.26 (WHO), depending on the growth reference used. Among adolescents with high baseline BMI z-score, the treatment difference varied almost 4-fold from -0.05 (CDC 2000) to -0.19 (WHO).

Visual inspection of the four growth charts reveals highly variable patterns in z-score reference curves with increasing z-score. Although most growth references are constructed using data from cross-sectional studies, they are routinely interpreted as growth trajectories in practice. For example, a child with a BMI z-score of 3.2 at one visit will be reevaluated against the same z-score with increasing age. This reevaluation may involve a qualitative

assessment using visual inspection of the child's BMI on a growth chart relative to percentile or z-score reference curves in routine clinical settings, or it may be quantified in the setting of observational studies or clinical studies. Therefore, the trajectory of the reference curve determines the interpretation of weight status at follow-up. Given that reference curves are used in this way, the observation that WHO and IOTF BMI z-score reference curves decrease with increasing age during adolescence is counterintuitive.

Differences among the growth references arise from their selection of reference population as well as the statistical methods^{9–12,15,16} used to describe BMI distributions in these populations, as summarized in Table 2. Values of BMI z-score at baseline, change in BMI z-score from baseline to week 56, and treatment differences were similar for the CDC 2000 and CDC Extended growth references in the lower baseline BMI z-score group because the two references are similar in this range, but results diverge in the higher baseline BMI z-score group (Fig. 2) because of differences in data source and methods at higher BMI z-score values.

There is no single “best” growth reference for every study, and the selection of growth reference for a given study depends on factors such as the BMI distribution and geographic origin of the study population as well as the intended use of study results. Investigators conducting a study in a single country will naturally use the official growth reference for that country. However, this practice makes it challenging to pool results from different countries. Studies conducted in multiple countries may use an international growth reference or select a single country's growth reference. Comparing studies that include children and adolescents with more severe obesity, where differences among the growth references are the most apparent, can be particularly challenging. Comparing results from diverse studies could be facilitated by reporting study results using BMI z-scores adopting multiple growth references.¹⁷ Finally, identifying the growth reference when citing z-score or percentile results is important for reducing inappropriate comparisons among different studies.

Conclusions

The results of this reanalysis of a weight management clinical study in pubertal adolescents with obesity illustrate the potential pitfalls of assuming comparability of BMI z-scores from different growth references, which occurs in meta-analyses and when z-scores are cited without the associated growth reference.¹ Despite all growth references producing the same overall conclusion of statistically significant weight loss, the magnitude of the treatment effect varied widely. Depending upon the reference population used, the BMI z-score can produce highly variable treatment estimates and alter interpretations of clinical meaningfulness. BMI z-scores cited without the associated growth reference cannot be accurately interpreted.

Funding Information

No external funding supported this research.

Data Sharing Statement

Data will be shared with bona fide researchers who submit a research proposal approved by an independent review board. Individual patient data will be shared in datasets in a de-identified and anonymized format. Data will be made available after research completion and approval of the product and product use in the European Union and the United States. Information about data access request proposals can be found at novonordisk-trials.com.

References

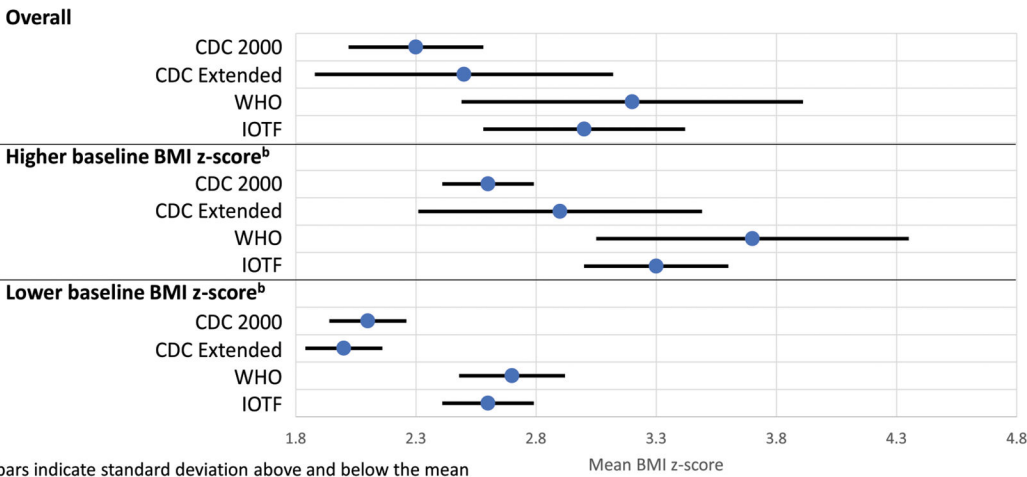
1. Grossman DC, Bibbins-Domingo K, Curry SJ, et al. Screening for obesity in children and adolescents: US preventive services task force recommendation statement. *Jama* 2017;317(23):2417–2426; doi: 10.1001/jama.2017.6803 [PubMed: 28632874]
2. Reinehr T, Lass N, Toschke C, et al. Which amount of BMI-SDS reduction is necessary to improve cardiovascular risk factors in overweight children? *J Clin Endocrinol Metab* 2016;101(8): 3171–3179; doi: 10.1210/jc.2016-1885 [PubMed: 27285295]
3. Ford AL, Hunt LP, Cooper A, Shield JP. What reduction in BMI SDS is required in obese adolescents to improve body composition and cardiometabolic health? *Arch Dis Child* 2010;95(4):256–261; doi: 10.1136/adc.2009.165340 [PubMed: 19966092]
4. El-Medany AYM, Birch L, Hunt LP, et al. What change in body mass index is required to improve cardiovascular outcomes in childhood and adolescent obesity through lifestyle interventions: A meta-regression. *Child Obes* 2020;16(7):449–478; doi: 10.1089/chi.2019.0286 [PubMed: 32780648]
5. O'Connor EA, Evans CV, Burda BU, et al. Screening for obesity and intervention for weight management in children and adolescents: Evidence report and systematic review for the US Preventive Services Task Force. *Jama* 2017;317(23):2427–2444; doi: 10.1001/jama.2017.0332 [PubMed: 28632873]
6. Food and Drug Administration. Guidance for Industry Developing Products for Weight Management. 2023. Available from: <https://www.fda.gov/media/71252/download> [Last accessed: July 5, 2023].
7. European Medicines Agency. Clinical evaluation of medicinal products used in weight control—addendum on weight control in children—scientific guideline. Available from: <https://www.ema.europa.eu/en/clinical-evaluation-medicinal-products-used-weight-control-addendum-weight-control-children>
8. Kelly AS, Auerbach P, Barrientos-Perez M, et al. A randomized, controlled trial of liraglutide for adolescents with obesity. *N Engl J Med* 2020;382(22):2117–2128; doi: 10.1056/NEJMoa1916038 [PubMed: 32233338]
9. Kuczmarski RJ, Ogden CL, Guo SS, et al. 2000 CDC growth charts for the United States: Methods and development. *Vital Health Stat* 11 2002;(246):1–190.
10. Hales CM, Freedman DS, Akinbami L, et al. 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 Health Stat* 1 2022;(197):1–42.
11. de Onis M, Onyango AW, Borghi E, et al. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ* 2007;85(9):660–667. [PubMed: 18026621]
12. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatr Obes* 2012;7(4):284–294; doi: 10.1111/j.2047-6310.2012.00064.x [PubMed: 22715120]
13. R: A Language and Environment for Statistical Computing. 2022.
14. Anthroplus: Computation of the WHO 2007 References for School-Age Children and Adolescents (5 to 19 Years). 2021.
15. World Health Organization. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.

2023. Available from: http://www.who.int/childgrowth/standards/technical_report/en/ [Last accessed: July 17, 2023].

16. World Health Organization. Computation of Centiles and Z-scores for Height-for-age, Weight-for-age, and BMI-for-age. 2023. Available from: https://cdn.who.int/media/docs/default-source/child-growth/growth-reference-5-19-years/computation.pdf?sfvrsn=c2ff6a95_4 [Last accessed: July 13, 2023].
17. Ryder JR, Kelly AS, Freedman DS. Metrics matter: Toward consensus reporting of BMI and weight-related outcomes in pediatric obesity clinical trials. *Obesity* (Silver Spring) 2022;30(3):571–572; doi: 10.1002/oby.23346 [PubMed: 35043581]

Impact Statement

Although BMI z-scores from different growth references are often considered comparable, reanalysis of a study of liraglutide in adolescents showed 2-fold differences in overall treatment effect and 4-fold differences among adolescents with higher baseline BMI. Our findings challenge current assumptions of BMI z-score comparability by highlighting discordant interpretations of clinical meaningfulness.



^a Error bars indicate standard deviation above and below the mean
^b Based on baseline BMI z-score (WHO) above or below 3.17.
BMI, body mass index; CDC, Centers for Disease Control and Prevention; IOTF, International Obesity Task Force; WHO, World Health Organization.

Figure 1. Mean BMI z-score at baseline, by high or low baseline BMI z-score. ^aError bars indicate standard deviation above and below the mean. ^bBased on baseline BMI z-score (WHO) above or below 3.17. IOTF, International Obesity Task Force; WHO, World Health Organization.

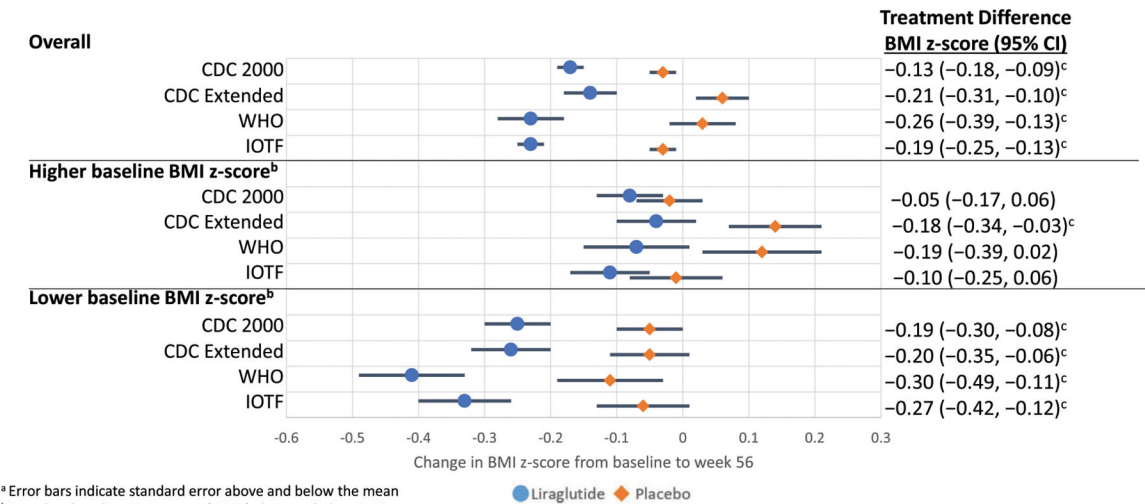
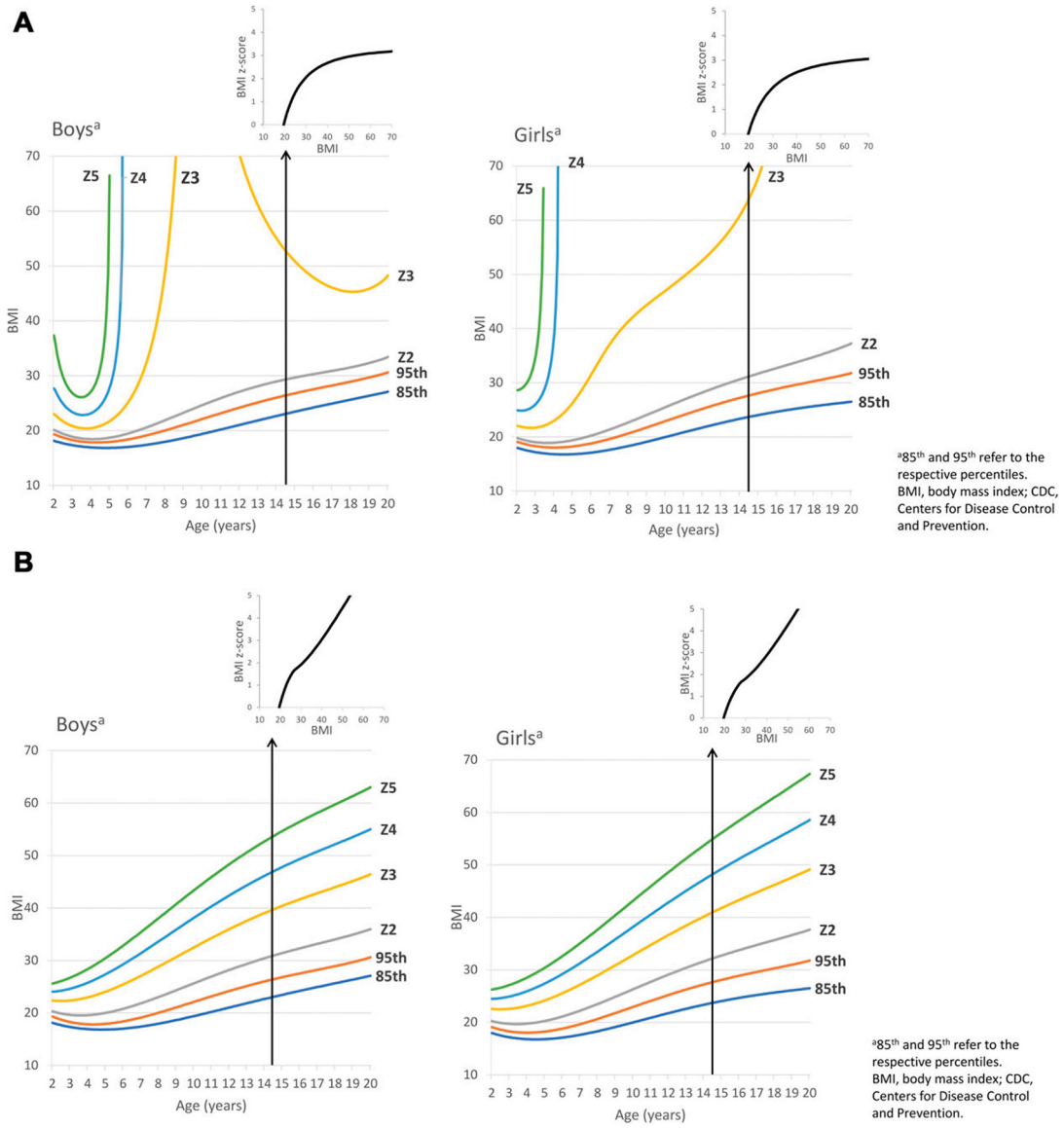
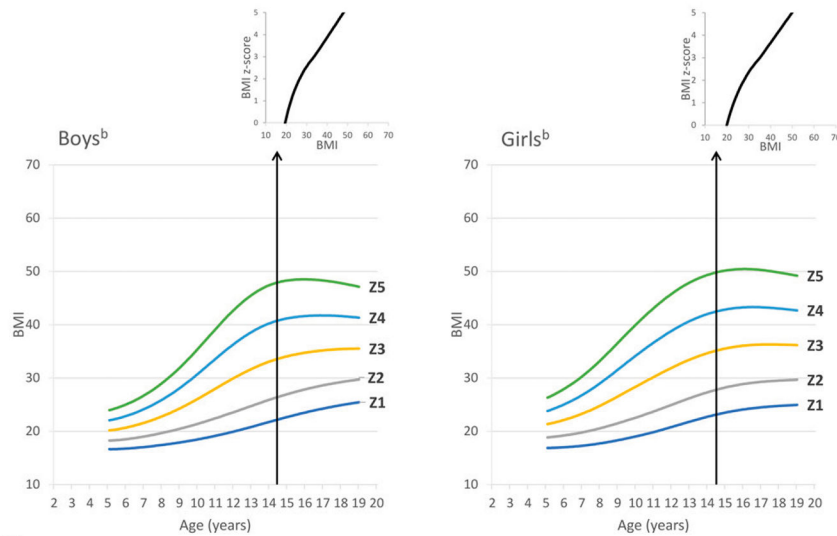
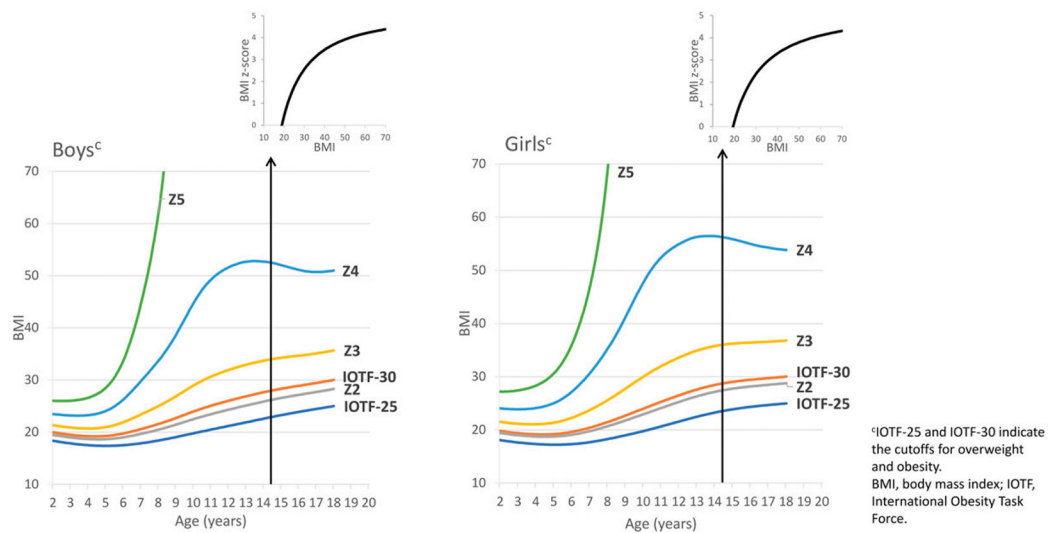


Figure 2. Change in BMI z-score from baseline to week 56 and treatment difference by baseline BMI z-score. Error bars indicate standard error above and below the mean. ^aBased on baseline BMI z-score (WHO) above or below 3.17. ^bStatistically significant (p < 0.05). CI, confidence interval.



C**D****Figure 3.**

BMI-for-age growth references with curves indicating BMI z-scores 1–5 and subplots of BMI z-score vs. BMI at age 14.5 years. (A) CDC 2000 growth reference. (B) CDC Extended growth reference. (C) WHO growth reference. (D) IOTF growth reference. ^a85th and 95th refer to the respective percentiles. ^bWHO cutoffs for overweight and obesity are the sex- and age-specific Z1 and Z2, respectively. ^cIOTF-25 and IOTF-30 indicate the cutoffs for overweight and obesity.

Table 1. BMI at Baseline and Week 56, Change in BMI from Baseline, and Treatment Difference, by High or Low Baseline BMI

	Treatment group	Sample size, n	Median BMI at baseline (IQR), kg/m ²	Mean BMI at baseline (SD), kg/m ²	Absolute BMI change from baseline (SE), kg/m ²	Relative BMI change from baseline (SE), %	Treatment difference (95% CI), kg/m ²
Overall	Liraglutide	125	34.4 (29.3, 44.3)	35.3 (5.1)	-1.4 (0.3)	-4.4 (0.8)	-1.8 (-2.6, -1.0)
	Placebo	126	34.1 (29.6, 48.4)	35.8 (5.7)	0.4 (0.3)	0.9 (0.8)	
Higher baseline BMI group ^a	Liraglutide	65	37.9 (33.4, 47.8)	38.7 (4.7)	-0.5 (0.5)	-1.6 (1.3)	-1.5 (-2.7, -0.2)
	Placebo	61	38.6 (33.7, 50.1)	40.1 (5.2)	1.0 (0.6)	2.2 (1.6)	
Lower baseline BMI group ^a	Liraglutide	60	31.5 (27.7, 35.1)	31.7 (2.2)	-2.4 (0.5)	-7.4 (1.4)	-1.9 (-3.1, -0.7)
	Placebo	65	31.7 (28.8, 34.5)	31.8 (1.8)	-0.5 (0.5)	-1.2 (1.4)	

^aBased on baseline BMI z-score (WHO) above or below 3.17.
CI, confidence interval; SD, standard deviation; SE, standard error; WHO, World Health Organization.

Table 2.
Summary of Data Sources and Methods for Constructing z-Scores for Four Growth References

Growth reference (age range)	Data sources	Methods for constructing z-scores	Cutoffs for overweight and obesity
CDC 2000 (2–20 years)	Nationally representative survey of US children and adolescents collected primarily from the 1960s through 1982	Modified LMS method for characterizing sex- and age-specific BMI distributions by their mean, dispersion, and skewness	85th percentile; 95th percentile
CDC Extended (2–20 years)	1. Nationally representative survey of US children and adolescents collected primarily from the 1960s through 1982 2. Children and adolescents with obesity from the CDC 2000 reference and US national surveys through 2016	Up to the 95th percentile: Same as the CDC 2000 growth reference Above the 95th percentile: Modeled using half-normal distributions for children and adolescents with obesity from the CDC 2000 reference and US national surveys through 2016	85th percentile; 95th percentile
WHO (5–19 years)	1. Nationally representative survey of US children and adolescents collected from the 1960s through 1974 (i.e., the 1977 National Center for Health Statistics reference population) 2. Children aged 18–71 months in the WHO Child Growth Standard for children aged 0–5 years	Outliers (such as those with BMI above +3 standard deviations of the median and under –3 standard deviations) were excluded to prevent the influence of “unhealthy” values. The BMI distributions were characterized using the LMS method, but only for the interval between z-scores –3 and 3. To calculate the BMI z-score for BMI values above Z3, an arbitrary sex- and age-specific standard deviation—set as the difference in BMI values between Z2 and Z3—was applied. Children aged 18–71 months in the WHO Child Growth Standard (for children aged 0–5 years) were also included to ensure a smooth transition between the two WHO growth references.	Z1; Z2
IOTF (2–18 years)	Nationally representative survey data from the United Kingdom, the United States, the Netherlands, Brazil, Singapore, and Hong Kong collected between 1963 and 1993	Country-specific LMS parameters were fit and then pooled to construct international percentile cutoffs for overweight and obesity. Cutoffs vary by sex and age and are calibrated such that they correspond to a BMI of 25 kg/m ² for overweight and 30 kg/m ² for obesity at age 18 years.	IOTF 25: 90.5th (boys) and 89.3rd (girls) percentiles; IOTF 30: 98.9th (boys) and 98.6th (girls) percentiles

IOTF, International Obesity Task Force.