Economic Research Service
U.S. DEPARTMENT OF AGRICULTURE

Economic Research Service

Technical Bulletin 1960

June 2023

Assessment and Adjustment of Body Weight Measures in Scanner Data

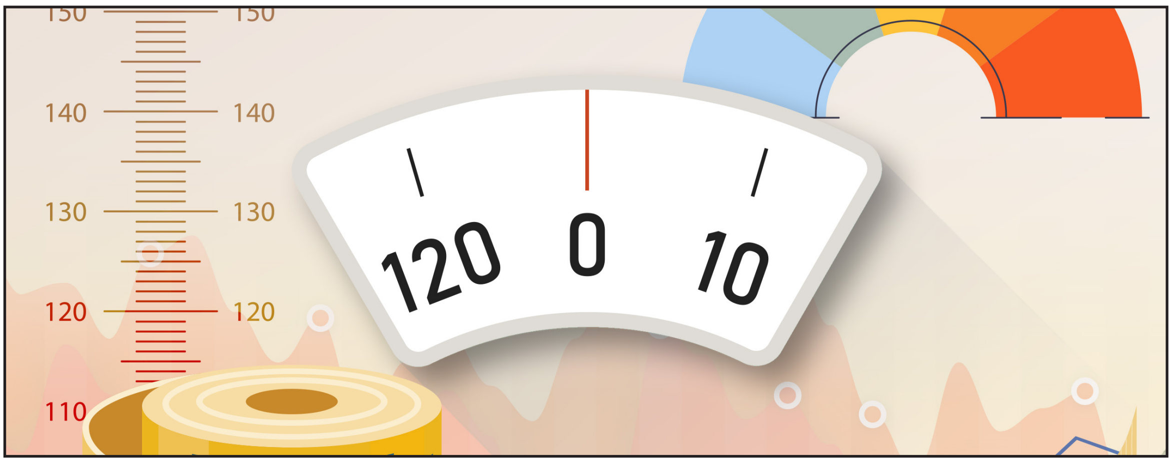
Sabrina K. Young, Elina T. Page, Abigail Okrent, Megan Sweitzer

Economic Research Service
U.S. DEPARTMENT OF AGRICULTURE

Economic Research Service
www.ers.usda.gov

Recommended citation format for this publication:

Young, Sabrina K., Elina T. Page, Abigail Okrent, and Megan Sweitzer, June 2023. *Assessment and Adjustment of Body Weight Measures in Scanner Data*, TB-1960, U.S. Department of Agriculture, Economic Research Service.



Cover is a derivative of images from Getty.

Use of commercial and trade names does not imply approval or constitute endorsement by USDA. The analysis, findings, and conclusions expressed in this report should not be attributed to Circana (formerly IRI).

To ensure the quality of its research reports and satisfy governmentwide standards, ERS requires that all research reports with substantively new material be reviewed by qualified technical research peers. This technical peer review process, coordinated by ERS' Peer Review Coordinating Council, allows experts who possess the technical background, perspective, and expertise to provide an objective and meaningful assessment of the output's substantive content and clarity of communication during the publication's review.

USDA Economic Research Service
U.S. DEPARTMENT OF AGRICULTURE

Economic Research Service

Technical Bulletin 1960

June 2023

Assessment and Adjustment of Body Weight Measures in Scanner Data

Sabrina K. Young, Elina T. Page, Abigail Okrent, and Megan Sweitzer

Abstract

Household scanner data are a rich resource for understanding food purchasing habits in the United States. The IRI Consumer Network provides a detailed account of the retail food purchases for a large, nationally representative sample. These data further include self-reported height and weight for a subset of households that complete the MedProfiler survey. Together, the Consumer Network and MedProfiler surveys provide a unique opportunity to study the relationship between diet and obesity. This report includes an assessment of the MedProfiler height and weight data in determining body mass index (BMI) for children and adults, using MedProfiler data from 2012 to 2018 and National Health and Nutrition Examination Survey data from 2011–2012 to 2017–2018. In addition, because self-reported height and weight may often be misreported in survey data, the report explores adjustment methods to account for any self-reporting measurement bias. Finally, since food-purchase data are collected at the household level, the report includes a comparison of methods for defining the obesity status of a household.

Keywords: IRI, Consumer Network, MedProfiler, body mass, BMI, obesity, scanner data, food expenditures, self-reporting bias, measurement bias

Acknowledgments

The authors would like to thank Annemarie Kuhns, previously with U.S. Department of Agriculture, Economic Research Service (USDA, ERS) for support in early iterations of this work, as well as Katherine Ralston, Debbie Rubas, Mark Prell, Guijing Wang, and Lisa Mancino of USDA, ERS for reviews and advice related to this report. They also thank the following individuals for technical peer reviews: Ricky Volpe, California Polytechnic State University, Edward Jaenicke, Pennsylvania State University, and Brandon Restrepo, ERS. Thanks also to ERS staff members Grant Wall, Elaine Symansky, and Courtney Knauth for editing the report and Jeremy Bell for the layout and design.

About the Authors

Sabrina K. Young, Elina T. Page, Abigail Okrent, and Megan Sweitzer are research economists at USDA, ERS.

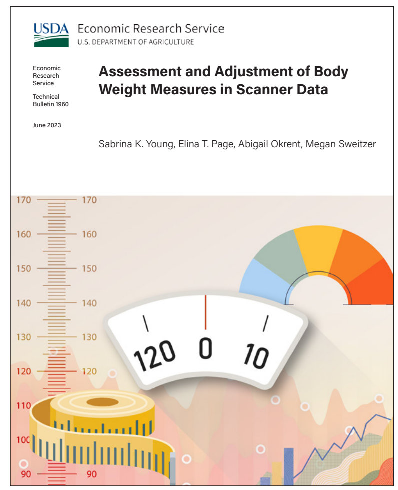
Contents

USDA Economic Research Service
U.S. DEPARTMENT OF AGRICULTURE

A report summary from the Economic Research Service *June 2023*

Assessment and Adjustment of Body Weight Measures in Scanner Data

Sabrina K. Young, Elina T. Page, Abigail Okrent, and Megan Sweitzer



What Is the Issue?

The high prevalence of obesity in the United States has health, social, and economic costs for both the affected individuals and society as a whole. The role that diet plays in obesity is an important area of research. A useful proxy for diet is available in scanner data from the IRI Consumer Network, which provides a weekly picture of household food-at-home purchases. A subset of households in the IRI Consumer Network—the IRI MedProfiler—also reports height and weight for each household member. This enables researchers to calculate body mass index (BMI) and investigate relationships between BMI and food purchases, an important link in the fight against obesity and chronic disease. However, self-reported height and weight are often misreported in survey data. Biases may be more pronounced for some demographic groups, such as those for age, gender, and/or race/ethnicity, increasing the risk of misrepresenting some groups more than others in obesity research.

The ERS researchers who authored the report compare self-reported BMI from the IRI MedProfiler to the measured BMI from the National Health and Nutrition Examination Surveys (NHANES) to assess the quality of the IRI data and develop methods for improving it for use in research. The study also examines different ways to define household body weight status, for example, obesity status of the primary shopper or of all household members.

What Did the Study Find?

BMI based on parent- and self-reported height and weight in the IRI MedProfiler differs from BMI based on measured height and weight in NHANES. For children and youth (ages 2 to 19):

- Average reported BMI in the MedProfiler (19.79 kg/m^2) is lower than measured BMI in NHANES (20.59 kg/m^2).
- Underweight (13 percent) and obese (20 percent) children and youths are more prevalent in the MedProfiler compared to NHANES (4 percent for underweight and 19 percent for obesity).
- Almost all distributions of BMI for children/youth between ages 2 and 19 by gender reported in the MedProfiler were statistically different from their measured counterparts in NHANES.

Although systematic differences exist in adult BMI distributions between the MedProfiler and NHANES, the differences are generally smaller than those noted for children and youths. For adults age 20 and older:

- Average reported BMI in the MedProfiler is lower (28.86 kg/m²) than measured BMI in NHANES (29.23 kg/m²).
- Obese adults are less prevalent in the MedProfiler (35 percent) than in NHANES (39 percent).

All distributions of BMI by race and ethnicity and gender reported in the MedProfiler are statistically different from their measured counterparts in NHANES. The study explored several methods for adjusting BMI distributions to reduce measurement bias because of self-reported BMI data from the MedProfiler. The only method that resulted in an improved prediction of measured BMI was based on a percentile-ranking regression model of self-reported BMI. This option is available just for adults, however, because NHANES collects only measured height and weight—not self- or parent-reported—for individuals under age 16. For children and youth, unadjusted data are preferred to outlier exclusion methods that the authors tested, which alter demographic characteristics of the sample but do not improve prediction.

The share of households classified as obese changed considerably with differing definitions of household obesity and across household characteristics. Differences were especially pronounced for larger household sizes and households with children.

How Was the Study Conducted?

This study examined patterns in body weight status across individuals and households using the IRI Consumer Network household panel survey and the IRI MedProfiler survey from 2012 to 2018. The ERS researchers used height and weight data from the National Health and Nutrition Examination Survey (NHANES) of the National Center for Health Statistics to compare with self-reported height and weight data from the MedProfiler. To correct for possible measurement bias in BMI calculations based on MedProfiler data, the researchers considered three adjustment methods: (1) removing outliers based on the minimum and maximum measured height and weight values reported in NHANES; (2) removing outliers based on the MedProfiler interquartile range; and (3) predicting BMI in the IRI MedProfiler using measured BMI and percentile rankings of self-reported BMI in NHANES.

Using the percentile-ranking adjustment method for adults, with no adjustments for children and youths, the researchers classified households by body weight status as normal weight, overweight, and obese. They compared household obesity levels for four possible ways of defining obesity at the household level, based on obesity of: (1) the primary shopper, (2) any member of the household, (3) at least half the household members, or (4) all household members. In a subsample of households with children, four additional definitions of household obesity were considered based on obesity of: (1) the primary shopper, (2) any child, (3) at least one adult and one child, or (4) all household members. Finally, the researchers compared results for household obesity status for all households, by race and ethnicity of the primary shopper, and—for all households only—by the number of household members.

Assessment and Adjustment of Body Weight Measures in Scanner Data

Introduction

The obesity epidemic is a global phenomenon, with the obesity rate nearly tripling since 1975 (World Health Organization, 2018). In the United States, the rate of obesity is higher than that of most countries, with over 40 percent of adults and about 21 percent of adolescent youths classified as obese (Ogden et al., 2020). By 2030, it is projected that nearly 50 percent of adults in the United States will be classified as obese, with nearly 25 percent of adults projected to have severe obesity (Ward et al., 2019).

Increasing U.S. obesity rates are linked to higher risks of morbidity and mortality from chronic illnesses attributable to excess fat accumulation, such as type 2 diabetes, cardiovascular disease, and high blood pressure. The economic impact of obesity on those affected directly by it and for the broader society was estimated to be roughly $1.7 trillion in 2016 dollars, equivalent to 9.3 percent of the U.S. gross domestic product (GDP) (Waters and Graf, 2018). This includes $480.7 billion in direct health care costs and an additional $1.24 trillion in indirect costs due to lost economic productivity.

Diet is the primary contributing factor to obesity and the leading risk factor associated with death in the United States (U.S. Burden of Disease Collaborators, 2018). Several studies have shown a link between body mass index (BMI), a standard measure of body fat based on height and weight, and consumer food choices. Diets low in essential nutrients and high in calories, with high consumption of sugar-sweetened beverages, sweets, refined cereals, solid fats, and red and processed meats, are associated with obesity (see Wirfält et al. (2013) for a review). In general, U.S. residents tend to consume more calories than they need, and the composition of foods they consume is not consistent with dietary guidelines (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020; Mancino et al., 2018).

Household scanner data, which include commercial data on household food purchases, are a rich source of information on food purchasing patterns. These data have been used to study food demand (Dong et al., 2018), food purchasing patterns across store formats and proximity (Volpe et al., 2017; Rahkovsky and Snyder, 2015), diet quality (Volpe and Okrent, 2012; Carlson et al., 2019), and the impact of specific food policies, such as taxing sugar-sweetened beverages (Zhen et al., 2014).

Scanner data also allow researchers to link household food purchases to self-reported height and weight of household members in order to better study and understand the relationship between diet and obesity. Because a subsample of households report health measures for all household members including height and weight, household scanner data can also be used to track changes in BMI and its association with specific types of foods purchased by households over time (Chen et al., 2019). However, self-reported height and weight are often misreported in survey data, and it is necessary to consider the quality of the data when calculating BMI and classifying household members as normal weight, overweight, and obese.

Studies show that adults tend to overreport height and underreport weight, and the degree of misreporting varies with age, body weight, gender, and other sociodemographic characteristics. Rowland (1990) found in self-reported and measured height and weight data from the National Health and Nutrition Examination Survey (NHANES) II (1976–80) that height and weight were reported on average with small errors, but larger systematic differences in the average reporting error were found in important population subgroups. In particular, heavier people underreport their weight more than lighter people, and older people overreport their height more than younger people. Kuczmarski et al. (2001) found age was an important factor in classifying body weight, height, BMI, and body weight status from self-reports in NHANES III (1988–94). They concluded that overestimation of height by older adults may occur because a longer time has elapsed since height was last measured and height decreases with age. Additionally, men are more likely than women to overestimate their height, while women, particularly young women, are more likely to underestimate their weight (Spencer et al., 2002; Bolton-Smith et al., 2000). Race and ethnicity were also found to affect self-reported height and weight (Stommel and Schoenborn, 2009; Wen and Kowaleski-Jones, 2012).

Using unadjusted self-reported height and weight in scanner data may cause bias in research on obesity and food consumption. Studies have found that use of self-reported height and weight have implications for food insecurity and mortality estimates by obesity status compared to using measured values (Lyons et al., 2011; Keith et al., 2011). Since biases may be larger for some demographic groups, adjustments have the potential to reduce bias in research due to differential misreporting.

The degree of misreporting can also vary with how a survey is administered. In-person interviews where the subject expects to have height and body weight measured in the future, as is done in NHANES, results in less self-reported mismeasurement compared to telephone- or web-based surveys, where there is no such expectation (Courtemanche et al., 2015).

Parent-reported height and weight for children may also suffer from measurement biases because children are continuously growing, making it difficult for parents to keep accurate measurements. This may lead to parental reports that underestimate the measured height of children, especially for children going through puberty (Wright et al., 2018). However, in a review of studies that made direct comparisons, the mean parent-reported child height and weight were close to the corresponding measures' means, usually within 1 centimeter or 1 kilogram (Himes, 2009). Himes noted some exceptions to this; specifically, Mexican-American mothers underestimated child height in the U.S. Hispanic Health and Nutrition Examination Survey.[1] Himes concluded that the prevalence of child obesity based on parental reports did not systematically differ from prevalence based on direct measurements.

The first objective of this study is to validate and develop adjustments for self-reported height and weight measures available in household scanner data. Specifically, the researchers compared self-reported height and weight from the MedProfiler to self-reported and measured height and weight in NHANES to better understand the quality of the IRI data. They then explored methods to correct for measurement bias for children and youths and for adults. These included two outlier-removal methods for both age groups. The researchers also predicted adult BMI in the IRI MedProfiler based on measured BMI and percentile rankings of self-reported BMI in NHANES. By ensuring a reliable measure of BMI in the MedProfiler data, the goal was to ultimately improve understanding of the differences in food purchasing patterns among households classified as normal weight, overweight, and obese.

Government survey datasets that include food consumption—most notably NHANES—are typically collected at the individual level rather than the household level. Scanner data, on the other hand, collect food consumption data (in the form of food-at-home purchases) at the household level. This is a challenge when using scanner data in obesity research since obesity is an individual characteristic and no global standard currently exists on how to define it at the household level. Therefore, an additional report objective is to consider methods for defining the obesity status of a household. The researchers compared several definitions based on the obesity status of household members and the presence of children. They compared household obesity rates for all households, by race and ethnicity of the primary shopper, and—for all households only—by the number of household members.

[1] Hispanic may be any race; race categories used in this report exclude those of Hispanic origin.

Data Description

The primary data used in this analysis were from the IRI Consumer Network household scanner and the IRI MedProfiler survey.[2] IRI Consumer Network household scanner data are derived from a nationwide panel of over 120,000 households each year. Participants provide a detailed account of what food products they purchased and when and where they shopped.[3] After households are recruited, they download a mobile application or are provided with a handheld scanner to scan or input all their food purchases and transmit their purchase data on a weekly basis via the internet. The household purchase data include product characteristics (e.g., brand, size, type) and some limited nutrition data reported on the Nutrition Facts Panel, which together give a robust picture of the types of foods that households purchased.[4] The household demographic data include standard characteristics (e.g., household size, income, education, and race). These data are initially collected when panelists are recruited, and all panelists are prompted to update their household demographic data annually each January.

The Consumer Network is a nonprobability sample in which households are selected for panel membership through stratified quota random sampling. Households are selected based on characteristics that best represent the U.S. population in the 48 contiguous States. Selection is based on meeting quotas based on demographic targets, such as household size, age of household head, race, ethnicity, education, occupation, presence of children, and area of residence (Muth et al., 2016).

To ensure data quality, IRI checks the consistency of weekly data reporting of panelists and identifies households in the final sample that consistently report purchases throughout the calendar year (also called the static panel). About one-half of the recruited households are included in the static panel and are assigned projection factors (i.e., survey weights).

In addition to households reporting products with a barcode or Universal Product Code (UPC), a growing subsample of the IRI static households also report purchases of random-weight products without a UPC (increasing from 54 percent of households in 2012 to 86 percent in 2018). These products are sold by the pound or count and include fresh fruits and vegetables, meat, cheese, baked goods, prepared foods, coffee, and bulk candy, nuts, and seeds. Households in this subsample are known as the random-weight panel. For the assessment of BMI based on self-reported height and weight, this study included only households that report both UPC and random-weight purchases, since those households submit a complete report of their food purchases and are more likely to be used in consumer food research.

The IRI MedProfiler is an opt-in survey on individual health and medical conditions offered to all households in the Consumer Network each October. Between 2012 and 2018, over 50 percent of the static panel that also reported random-weight purchases had at least one member respond to the MedProfiler survey in a given year (ranging from 17,072 to 30,784 households), with responses received from 40,118 to 69,713 individuals. In the survey, adults 18 years or older are asked to report height and weight for themselves and for children in the household. Individuals missing height or weight in the MedProfiler were excluded (1,188 individual observations), as were children under age 2 (3,022 child observations) in order to match the NHANES sample. In this analysis, each household-year and individual-year was treated as a single, unique observation.

The Consumer Network static sample includes post-stratification weights (also called projection factors) that

[2] Previous research by the U.S. Department of Agriculture examined the survey methodology, the representativeness of the demographic makeup, and the reported expenditures of the household panel. See Muth et al. (2016) and Sweitzer et al. (2017) for more information.

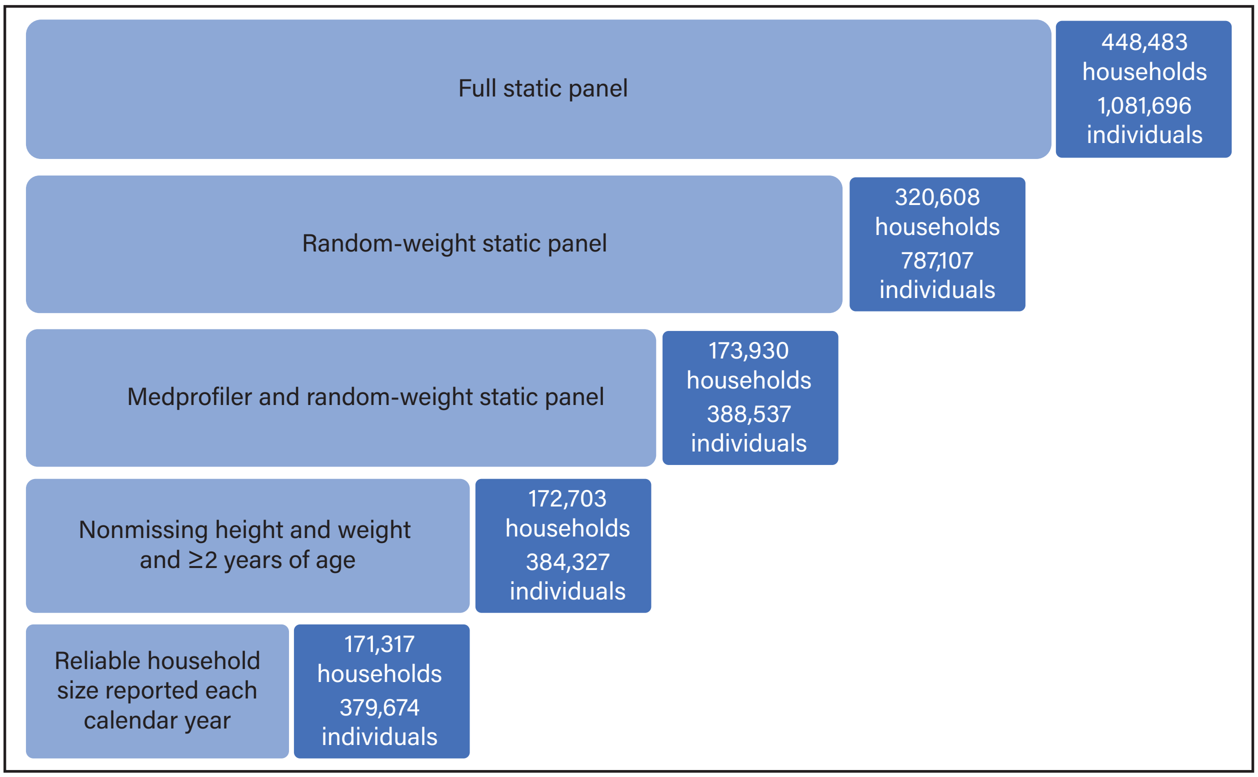
[3] The same household may participate in more than 1 year.

[4] More comprehensive nutrition data for food purchases reported in the IRI scanner data are available through the ERS Purchase to Plate Crosswalk (Carlson et al., 2022).

weight the data to match U.S. Census demographic targets, which help account for differences between the composition of the static panel and the U.S. general population. Households included in the MedProfiler are assigned a separate set of projection factors to weight the MedProfiler subsample to the U.S. population, and these were used in all calculations throughout this analysis.

Because household demographic data and MedProfiler data are collected at different times of the year—January and October, respectively—differences in reported household composition are expected. For example, a household in January may report a household size of two in the Consumer Panel sample and then a household size of three in the MedProfiler sample in October of that year because of a birth. Small changes in household size can plausibly occur between January and October because of births, deaths, divorce, and so on. However, some changes in household size were deemed implausible for the analysis and assumed to be measurement error or a substantial change to household circumstances. To ensure quality of data reporting, 4,473 individuals (based on MedProfiler counts) for which the reported household size in the demographic file and the number of individuals responding to the MedProfiler survey differed by more than two were also excluded.[5] The final MedProfiler sample includes 320,682 adult observations (aged 20 and older) and 58,992 child and youth observations (aged 2 to 19).[6] Figure 1 outlines the MedProfiler sample refinement.

Figure 1
Sample refinements and unweighted sample sizes of the IRI Consumer Network, 2012–18



Note: Households and individuals that are in multiple years were counted as one unique observation in the cross section. Reliable household size refers to differences of no more than two between the demographic file and the number of individuals responding to the MedProfiler survey.

Source: USDA, Economic Research Service calculations based on the 2012–2018 IRI Consumer Network and MedProfiler Survey.

[5] Even if large differences in household size represent actual changes in household composition, the authors believe these households still warrant exclusion from analysis because the food purchasing behaviors of these households likely substantially changed within the reporting year.

[6] Since the projection factors scale the IRI MedProfiler-random weight sample to be consistent with demographic targets in the U.S. population for a year, the authors divided the projection factors by a factor of 7 to produce 7 year estimates over all cycles of the IRI MedProfiler.

Since evidence suggests that self-reported weight suffers from downward biases and height from upward biases, the researchers compared BMI based on self-reported height and weight in the MedProfiler to the BMI reported and measured in the continuous NHANES for the 2011–2012, 2013–2014, 2015–2016, and 2017–2018 cycles (2011–2018 in total).[7] [8] The NHANES surveys use complex, stratified, multistage probability sampling to best represent the noninstitutionalized U.S. civilian population. Both measured and self-reported height and weight information are collected in NHANES for the same individuals. Self-reported height and weight are first collected during in-person interviews, and measured height and weight are collected several weeks later in physical examination centers by trained health technicians. Measurements by trained technicians provide objective measures of height and weight that can be used to validate self-reported height and weight data (Connor Gorber et al., 2007; Courtemanche et al., 2015; and Flegal et al., 2019).

The final 2011–2018 NHANES sample consisted of 20,409 adults aged 20 years and older (with ages 80 and older in 1 category) and 13,080 children and youths aged 2 to 19 years. To achieve the final NHANES sample, children younger than 2 years (2,568 child observations) and pregnant women (247 individual observations) were excluded because BMI is not a reliable measure of body fat for these groups. Of this restricted sample, participants missing measured height or weight variables were also excluded (2,168 individual observations). Individuals aged 16 years and older were asked to report height and weight in interviews; however, these self-reported values were used only for adults aged 20 and older in order to appropriately calculate BMI by age for youth aged 16 to 19 rather than grouping them with adults. The NHANES sample was further restricted to exclude 684 adults aged 20 and older because they refused, didn't know, or were missing self-reported height or weight information. The sample weights for the examination component were used in the estimation.[9]

Methods

This section describes methods for assessing reporting error in BMI in the MedProfiler based on reported (self-reported and parent-reported) height and weight. This is done by comparing the distribution of BMI for select subpopulations in MedProfiler against measured BMI (i.e., constructed from measured height and weight) in NHANES. Because of systematic differences between the distributions of BMI reported in the two datasets, the researchers also described methods to correct for bias arising from BMI based on reported measures of height and weight in the MedProfiler.

Assessment of Reporting Error in BMI (kg/m2) in Household Scanner Data

Many studies using self-reported height and weight to construct BMI found systematic differences between those reported and actual measures (e.g., Courtemanche et al., 2015; Ezzati et al., 2006). Courtemanche et al. compared densities of reported height and weight in the Behavioral Risk Factor Surveillance System (BRFSS) and the American Time Use Survey (ATUS) with measured height and weight in NHANES—the "gold standard" (Centers for Disease Control and Prevention, National Center for Health Statistics). Using the Kolmogorov-Smirnov (KS) tests of equality between distributions, Courtemanche et al. found the BMI distributions based on NHANES were not equal to those of BRFSS and ATUS. Ezzati et al. (2006) compared obesity

[7] BMI is calculated as body mass (weight) in kilograms divided by the square of body height in meters and is applicable to the population aged 2 and older.

[8] NHANES data are available in 2 year cycles. The authors used NHANES 2011–2012 but only MedProfiler for 2012 because the USDA, ERS acquisition of IRI Consumer Network does not contain separate household demographic data for 2011 and the researchers preferred more precise demographic information for the analysis.

[9] The 2 year sample weights were combined to produce 8 year estimates over three cycles of NHANES by dividing the 2 year weights by a factor of 4 (Chen et al., 2018).

prevalence and average BMI based on measured BMI by age-gender cohorts in NHANES to their reported counterparts in BRFSS over time and across States. On average, they found that BMI reported in BRFSS was lower than BMI measured in NHANES in-person interviews across all age cohorts, but the magnitude of difference was larger for females.

Following both Courtemanche et al. and Ezzati et al., this study assessed the reporting error in BMI calculated from self- and parent-reported height and weight in MedProfiler by comparing its distribution and population-level obesity rates to measured BMI and obesity rates from NHANES. This analysis was conducted separately for children and youths aged 2 to 19 years and adults aged 20 years and older because children and youths are continuously growing, and overweight and obese cutoffs vary by age and gender. For adults, the BMI distributions within each dataset were examined by gender and demographic profile.[10] For both youths and adults, the demographics, average weight, height, BMI, and population-level prevalence of underweight, normal weight, overweight, and obese were compared across datasets. T-tests were used to determine if differences in demographic and body weight composition between the two samples were statistically significant.

In addition, the researchers used the two-sample Kolmogorov-Smirnov test (KS test) to determine if the distribution of BMI based on self-reported height and weight (MedProfiler) is drawn from the same distribution as the distribution of BMI based on measured height and weight (NHANES) (Massey, 1951). The test statistic, or D statistic, is formed under the null hypothesis that the samples were drawn from the same distribution:

$$D_{m,n} = \max_x |F_m(x) - F_n(x)|,$$

where x is the variable of interest, in this case BMI, and $F_m(x)$ and $F_n(x)$ are weighted empirical cumulative distribution functions of BMI based on NHANES (m) and MedProfiler (n), respectively.[11] Hence, a rejection of the null hypothesis indicates that the two variables were not drawn from the same distribution.

Adjustments for Reducing Measurement Bias

The researchers examined several ways to adjust the IRI MedProfiler BMI distributions to reduce measurement bias introduced from using reported rather than measured height and weight in BMI calculations. First, the researchers adjusted BMI by eliminating outliers for the child and youth distributions and the adult distributions. Outliers were removed from the full stacked IRI MedProfiler 2012–18 data in two ways: (1) based on the minimum and maximum measured height and weight values reported in NHANES (NHANES-outlier method) (e.g., Freedman et al., 2015) and (2) based on the interquartile range (IQR-outlier method) (e.g., Rousseeuw and Croux, 1993). The NHANES-outlier method limits the sample by excluding individuals outside of the maximum and minimum height and weight in NHANES, stratified by gender and race/ethnicity for adults and stratified by gender and age for children.[12] For example, the measured heights for adult White females in NHANES ranged from 53 inches (1.3 meters) to 74.5 inches (1.9 meters). The measured weights ranged from 71.4 pounds (32.4 kilograms) to 450.6 pounds (204.4 kilograms). IRI respondents with a reported height or weight outside either range were excluded as an outlier. The IQR-outlier method used the interquartile range (i.e., the range between the 25th and 75th percentile) to exclude heights and weights of individuals in each gender and race/ethnicity (or age for children and youths) that were more than 1.5 times the interquartile range below the 25th percentile or 1.5 times the interquartile range above the 75th percentile.

For adults aged 20 and older, the researchers corrected for BMI measurement bias present in the MedProfiler

[10] The count of children in each gender-age cell in NHANES was too small to extend the child and youth analysis by demographic profiles.

[11] The KS tests were performed using the user-written ksmirnov2 command in Stata (Mittag, 2012).

[12] The NHANES outlier method is applied to each gender-age cohort in the MedProfiler because children and youth are growing in these years.

using both self-reported and measured BMI in NHANES. This method is only possible for adults because NHANES only collects measured height and weight for individuals under age 16 (not self-reported); therefore, these adjustments for children and youths could not be used. In practice, the standard method for correcting self-reported height and weight measurement error is to regress measured height and weight on self-reported height and weight using a validation sample that contains both measures (e.g., NHANES). Then the estimated coefficients from this regression are used to predict the measured values in the survey of interest (e.g., MedProfiler). This adjustment has been used on self-reported height and weight in the National Longitudinal Survey of Youth (NLSY) (Cawley, 2004; Baum, 2009), Behavioral Risk Factor Surveillance System (BRFSS) (Chou et al., 2004; Ezzati et al., 2006), and National Population Health Survey (Canada) (Gotay et al., 2013; Larose et al., 2016). This standard method assumes that the validation survey and survey of interest are derived using the same measurement methods. As noted by Han et al. (2009) and Pinkston (2015), when NHANES respondents reported height and weight in person they were aware that in the future their height and weight would be measured in person. Both studies highlight that this is not the case with NLSY, in which respondents reported height and weight over the phone with no expectations of in-person measurement. The MedProfiler presents a similar concern.

Instead, the researchers used the Courtemanche et al. (2015) BMI adjustment method, which relies on weaker assumptions about the relationship between measured and reported values in the validation sample and the sample of interest. Rather than using reported values, this method predicts measured values using the percentile rank of the self-reported values in their respective distributions. Hence, the method is robust to differences across samples in the severity (or type) of measurement error—as long as the rankings of respondents based on reported values align with the rankings based on measured values in both the validation sample and the sample of interest (and that both samples represent the same population, e.g., nationally representative).

The steps to adjust for self-reporting bias using the Courtemanche et al. (2015) correction are as follows:

1. Estimate percentile rank of self-reported BMI in the validation dataset.[13]
2. Generate cubic basis splines (b-splines) of the percentile ranks from step 1, essentially splitting up the polynomial into segments.[14]
3. Regress measured BMI on age polynomials, and the generated splines from step 2 with the validation dataset for each race/ethnicity and gender group.
4. Predict measured BMI using the estimated coefficients from step 3 with the dataset of interest.

In sum, the researchers calculated two adjustment methods for children and youths (the two outlier methods) and three adjustment methods for adults (the outlier methods and the percentile ranking regression method). Using t-tests and KS tests, the researchers then evaluated whether the adjustment methods correct for self-reporting bias in the MedProfiler survey.

[13] While the steps outlined here adjust BMI directly, this method can also be applied to adjust self-reported height and weight.

[14] Because percentile ranks are distributed between 0 and 1, while actual (and reported) measures are not, regressing actual measures on simple polynomials of the percentile rank likely result in predicted values that are poor fits for actual measures. Instead, the authors generated cubic basis splines (b-splines) of the percentile ranks and regressed the measured values on b-splines in percentile ranks. The b-splines were estimated using the user-written b-spline command in Stata (Newsom, 2000).

Results for Children and Youths

The demographics, body weight class, and BMI distributions for children and youths aged 2 to 19 years based on measured values in NHANES were compared with those reported in the MedProfiler.[15] Because statistically significant differences were found between the measured and reported distributions, the authors also compared these distributions after excluding outliers based on minimum and maximum height and weight measured in NHANES and based on the MedProfiler IQR. Ultimately, the outlier methods applied to the child and youth MedProfiler sample did little to correct for measurement bias from using reported values in calculating BMI.

Validation of BMI in Household Scanner Data

The demographic and body weight class composition of the child and youth samples between NHANES and MedProfiler differ considerably (columns 1 and 2 of table 1). The NHANES child and youth sample was 52 percent White, whereas the MedProfiler sample was 64 percent White. About 24 and 14 percent of the NHANES child and youth sample was Hispanic and Black, respectively, compared to 18 and 11 percent of the MedProfiler sample. The difference in gender composition between the two datasets was not statistically significant and less than 1 percentage point in magnitude.

Table 1
Mean BMI (kg/m^2) and distributions of children and youths in NHANES and MedProfiler by demographic group, body weight class, and adjustment method

	(1)	(2)	(3)	
			Adjusted MedProfiler (reported)	
	NHANES (measured)	Unadjusted MedProfiler (reported)	NHANES outliers excluded	IQR outliers excluded
Sample size	13,080	58,992	57,805	58,319
Distribution of population by demographic group (percent)				
Hispanic	23.80 (1.83)	18.18*** (0.16)	17.91*** (0.16)	17.83*** (0.16)
White	51.98 (2.38)	63.80*** (0.20)	64.56*** (0.21)	64.59*** (0.20)
Black	13.99 (1.28)	11.05*** (0.13)	10.76*** (0.13)	10.59*** (0.13)
Asian	4.76 (0.47)	4.63 (0.09)	4.53 (0.09)	4.72 (0.09)
Other race or ethnicity	5.48 (0.37)	2.33*** (0.06)	2.24*** (0.06)	2.27*** (0.06)
Male	50.99 (0.61)	51.75 (0.21)	51.89** (0.22)	51.93** (0.21)
Mean BMI (kg/m^2)	20.59 (0.08)	19.79*** (0.03)	19.52*** (0.022)	19.13*** (0.020)

continued on next page ►

[15] For the MedProfiler, height and weight are reported by parents for children and youths under the age of 18 and self-reported for youths aged 18 and 19. BMI based on parent- or self-reported measures are referred to as reported throughout this section.

◄ continued from previous page

	(1)	(2)	(3)	
			Adjusted MedProfiler (reported)	
	NHANES (measured)	Unadjusted MedProfiler (reported)	NHANES outliers excluded	IQR outliers excluded
Distribution of population by body weight class (percent)				
Underweight	3.64 (0.27)	12.71*** (0.14)	11.66*** (0.14)	12.98*** (0.14)
Normal weight	62.39 (0.62)	53.90*** (0.21)	57.82*** (0.21)	57.01*** (0.21)
Overweight	15.50 (0.39)	13.25*** (0.14)	13.83*** (0.15)	13.73*** (0.15)
Obese	18.48 (0.55)	20.14*** (0.17)	16.69*** (0.16)	16.27*** (0.16)

BMI = body mass index

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. MedProfiler body weight data are reported by parents for children and youths under age 18 and self-reported for youths ages 18 and 19. Sample weights and projection factors were used in all calculations. For children and youths, body weight classification is based on age and gender (see Centers for Disease Control and Prevention, National Center for Health Statistics, 2021). The NHANES-outlier method excludes height and weight values that fall below or above the minimum and maximum height and weight for children and youths by age and gender reported in NHANES. The IQR-outlier method excludes height and weight values that fall below or above the interquartile range (IQR) for children and youths by age and gender. Asterisk (*), double asterisk (**), and triple asterisk (***) indicate that the t-test of a difference between the NHANES (measured) sample and the MedProfiler samples is significant at the 10-, 5-, and 1-percent levels, respectively.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–18 IRI MedProfiler Survey.

BMI based on measured height and weight for children and youths in NHANES is higher (20.6 kg/m^2) on average compared to its reported counterpart in the MedProfiler (19.8 kg/m^2). Children and youths in the MedProfiler are more likely to be reported as underweight (12.7 percent vs. 3.6 percent) or obese (20.1 percent vs. 18.5 percent) compared to children and youths in NHANES; they are less likely to be normal weight (53.9 percent vs. 62.4 percent) or overweight (13.3 percent vs. 15.5 percent). These differences are all statistically significant at the 1-percent level.

Differences in the measured and reported BMI distributions between the two datasets also vary by age and gender for the child and youth samples as shown in table 2. Except for a few ages, for both male and female children and youths, reported median BMI in the MedProfiler is generally lower than NHANES. Additionally, for both male and female children and youths, the 90th percentile is consistently lower in the MedProfiler starting at age 9. The differences in medians of each age-gender distribution between the measured NHANES and the reported MedProfiler are less than or equal to 1 kg/m^2 with the exception of male children and youths aged 14 and 18 and female children and youths aged 12, 14, and 17. However, the differences in the 90th percentiles are sometimes quite large, ranging from -3.9 kg/m^2 to 8.6 kg/m^2. Examining other moments of the distributions (table A.1), reported BMIs based on the MedProfiler sample were more skewed to the right, with fatter tails compared to distributions using NHANES across most of the age-gender groups. Consistent with these noted differences, the KS tests for each gender-age distribution indicated that the reported BMIs in MedProfiler and measured BMIs in NHANES were not sampled from the same population (table 2). The only exception to this was for males at age 12.

Table 2

Mean, percentiles, and tests of distributions of measured (NHANES) and reported (MedProfiler) BMI (kg/m^2) for children and youths by age and gender

	Mean		Percentiles				
			50th		90th		
Age and Sex	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)	KS test (D-statistic)
Male							
2	16.85	19.10	16.70	17.10	18.60	27.00	0.28***
3	16.58	17.50	16.20	16.27	18.30	21.94	0.21***
4	16.30	16.95	15.90	16.10	18.40	21.09	0.19***
5	16.43	16.99	16.10	15.95	18.60	21.74	0.18***
6	16.71	17.00	16.00	15.75	19.70	22.56	0.22***
7	17.35	17.09	16.40	16.10	22.20	22.59	0.18***
8	17.95	17.97	16.90	16.78	22.90	24.21	0.15***
9	18.50	18.32	17.40	17.43	24.40	23.80	0.12***
10	19.96	19.15	19.00	18.16	26.10	25.02	0.11***
11	20.89	19.89	19.70	18.82	27.10	25.10	0.14***
12	20.65	20.42	19.30	19.53	27.70	26.90	0.07
13	22.47	21.00	20.60	20.02	30.90	27.43	0.11***
14	22.99	21.92	21.60	20.52	30.20	28.24	0.13***
15	23.95	22.31	22.30	21.41	32.50	29.02	0.15***
16	24.22	23.01	22.80	21.94	31.40	30.03	0.12***
17	24.75	23.51	23.40	22.36	32.90	30.84	0.12***
18	25.64	23.94	24.10	22.95	34.20	30.89	0.12***
19	25.73	24.72	24.30	23.49	34.20	32.98	0.13***
Female							
2	16.47	18.89	16.40	16.78	18.20	26.85	0.32***
3	16.16	17.52	15.90	16.27	17.90	22.65	0.23***
4	16.41	16.83	16.10	15.94	18.70	21.70	0.18***
5	16.29	16.71	15.80	15.94	18.90	21.79	0.24***
6	16.85	16.94	16.00	15.94	21.00	21.70	0.17***
7	17.24	17.17	16.40	16.17	21.90	22.52	0.15***
8	18.19	17.83	17.20	16.77	23.20	23.95	0.17***
9	19.12	18.43	17.80	17.57	25.00	24.23	0.17***
10	19.68	19.11	19.10	18.20	25.30	24.70	0.10***
11	20.95	20.25	19.50	18.98	28.40	26.93	0.09***
12	22.37	20.68	21.30	19.79	29.50	26.57	0.18***
13	22.54	21.53	20.70	20.54	30.30	28.24	0.11***
14	23.62	22.44	22.40	21.14	30.10	29.75	0.14***
15	24.17	23.17	22.60	21.64	32.60	30.90	0.11***
16	24.92	23.52	22.90	21.95	33.40	31.63	0.10***
17	25.34	23.97	23.60	22.31	33.10	31.95	0.13***
18	25.99	24.55	23.70	22.80	36.80	32.92	0.11***
19	26.39	24.85	23.80	22.86	37.70	33.81	0.10**

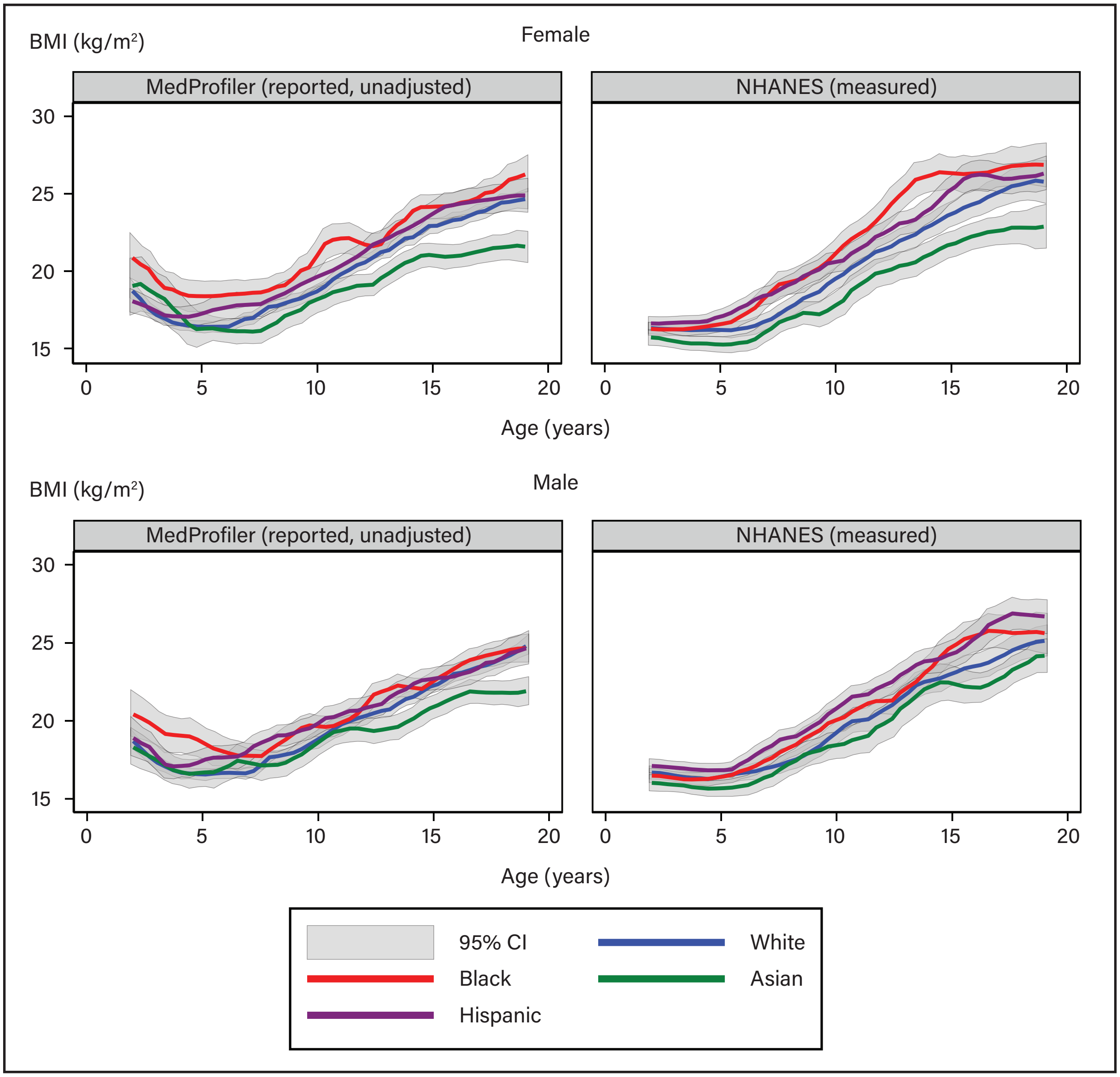
BMI = body mass index.

Note: MedProfiler body weight data are reported by parents for children and youths under age 18 and self-reported for youths ages 18 and 19. Sample weights and projection factors are used in all calculations. The Kolmogorov–Smirnov (KS) test compares the distributions for NHANES and MedProfiler datasets for each age-gender group. Asterisk (*), double asterisk (**), and triple asterisk (***) indicate that the D-statistic is significant at the 10-, 5-, and 1-percent levels, respectively.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey.

The overall relationship between age and BMI for children and youths based on reported values in the MedProfiler and measured values in NHANES is similar in terms of relative positioning of the estimated kernel-weighted local polynomial curves across demographic groups (figure 2). Generally, after age 5, BMI increases with age, with Asian children and youths having the lowest BMI, followed by White children and youths. However, notable differences also exist between the two samples in terms of the shape of relationship between age and BMI. In particular, BMI for children 8 years of age or less based on reported values in MedProfiler are markedly less than BMI based on measured values in NHANES. This produces a more pronounced J-shape relationship between age and BMI in the MedProfiler compared to NHANES. This J-shape is, however, in line with BMI-for-age percentiles published by the Centers for Disease Control and Prevention (Centers for Disease Control and Prevention, National Center for Health Statistics, 2017). In contrast, for children older than age 8, BMI based on MedProfiler tends to be less than that based on measured values NHANES. Hence, the slope of the relationship between age and BMI is greater for NHANES at older ages compared to the MedProfiler.

Figure 2
Relationship between age and BMI (kg/m^2) for children and youths by gender and demographic group



BMI = body mass index.

95% CI = 95-percent confidence interval. This measure indicates that there is 95 percent confidence that the true population parameter resides in this range.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors used in all calculations. Two implausible outliers (BMI>500) based on the IRI MedProfiler were excluded for ease of viewing overall trends.

Source: USDA, Economic Research Service calculations based on kernel-weighted local polynomial smoothing and the 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey data.

Adjustments for Reducing Measurement Bias

For children and youths, the researchers tested two methods for adjusting the sample to reduce measurement bias (column 3 of table 1). The first method limits the MedProfiler sample to children and youths within the minimum and maximum measured height and weight by gender and age in NHANES, which excludes 4,986 children and youths. The second method limits children and youths with height and weight values within the interquartile (IQR) range within each age and gender group, which excludes 3,940 children and youths.

Application of the outlier methods to the MedProfiler did not improve distributions of reported BMI for gender-age groups compared to NHANES. Both the IQR- and NHANES-outlier methods reduce average reported BMI for children and youths in the MedProfiler from 19.8 to 19.5 and 19.1 kg/m^2, respectively, which is still almost 1 kg/m^2 less than measured BMI (20.6 kg/m^2) in NHANES. Examining the adjustments by gender and age, the NHANES-outlier method mostly excludes children less than 6 years of age, so that the mean and 90th percentiles for these ages, especially for females, are markedly different from the unadjusted MedProfiler (table A.2).[16] In contrast, the IQR-outlier method mostly excludes youths 10 years of age and above, resulting in larger differences in mean and 90th percentiles between the unadjusted and IQR-adjusted samples at older ages.[17] The NHANES-outlier method also adjusts the skewness (which measures symmetry of a distribution) and kurtosis (which measures the thickness of a distribution's tails) of BMI distributions for most ages to better align with their measured counterparts, whereas the IQR-outlier method does little to adjust these upper moments (tables A.1 and A.2). Last, absolute differences between mean and median BMI across most of the age-gender groups based on the adjusted MedProfiler and the measured NHANES samples are larger than those based on the unadjusted MedProfiler and measured NHANES samples. Hence, in many cases, the outlier adjustment methods exacerbate differences between the reported BMI in the MedProfiler and those measured in NHANES.

Statistically significant differences also persist between NHANES and the adjusted MedProfiler samples with regard to the distribution of population by body weight class (table 1). The proportion of adjusted samples categorized as underweight is still significantly larger compared to those in the measured NHANES sample (4 percent)—12 percent for the NHANES outlier method and 13 percent for the IQR outlier method. Although the proportion of the MedProfiler sample categorized as normal weight, overweight, and obese modestly improved using both adjustment methods, the differences were still statistically significant compared with measured NHANES (table 1 and see also table A.3). Overall, these outlier methods do little to adjust the distributions of BMI based on reported values in the MedProfiler to align with those based on measured values in NHANES.

Results for Adults

As in the child and youth analysis, statistically significant differences were found between the measured and reported distributions in the adult sample, although the differences were smaller. In addition to excluding outliers to correct for measurement bias in the MedProfiler, the researchers also corrected for self-reporting bias using a percentile ranking regression.

[16] Of all excluded children and youths using the NHANES-outlier method, 51 percent are less than 6 years of age.

[17] Of all excluded children and youths using the IQR-outlier method, 57 percent are above the age of 10.

Validation of BMI in Household Scanner Data

Statistically significant differences exist between the MedProfiler and NHANES samples in terms of the demographic and body weight composition of adults (columns 1 and 2 of table 3). Compared to NHANES, the adult population of the MedProfiler is more White (70 percent versus 66 percent) and less male (45 percent versus 49 percent).

Table 3

Mean BMI (kg/m²) and distributions of adults in NHANES and MedProfiler by demographic group, body weight class, and adjustment method

	(1)		(2)	(3)		
				Adjusted MedProfiler (reported)		
	NHANES		Unadjusted MedProfiler (reported)	NHANES outliers excluded	IQR outliers excluded	Predicted percentile rank
Sample size	20,409		320,682	318,400	303,235	320,682
Distribution of population by demographic group (percent)						
Hispanic	14.00 (1.11)		13.05*** (0.06)	13.02*** (0.06)	13.17*** (0.06)	13.05*** (0.06)
White	65.62 (1.62)		70.06*** (0.08)	70.18*** (0.08)	70.04*** (0.08)	70.06*** (0.08)
Black	11.41 (0.96)		10.51*** (0.05)	10.51*** (0.05)	10.41*** (0.06)	10.51*** (0.05)
Asian	5.55 (0.48)		4.11*** (0.04)	4.09*** (0.04)	4.11*** (0.04)	4.11*** (0.04)
Other race or ethnicity	3.43 (0.24)		2.28*** (0.03)	2.21*** (0.03)	2.28*** (0.03)	2.28*** (0.03)
Male	48.89 (0.37)		44.92*** (0.09)	44.94*** (0.09)	45.09*** (0.09)	44.92*** (0.09)
	Measured	Reported				
Mean BMI (kg/m²)	29.23 (0.12)	28.49*** (0.11)	28.86*** (0.013)	28.81*** (0.01)	28.18*** (0.01)	29.27 (0.01)
Distribution of population by body weight class (percent)						
Underweight	1.53 (0.10)	1.57*** (0.12)	1.94*** (0.02)	1.83*** (0.02)	1.88*** (0.02)	1.08*** (0.02)
Normal weight	27.43 (0.64)	31.00*** (0.69)	29.89*** (0.08)	29.99*** (0.08)	30.94*** (0.08)	27.86 (0.08)
Overweight	32.33 (0.54)	33.52*** (0.56)	33.17 (0.08)	33.32*** (0.08)	34.48*** (0.09)	32.42 (0.08)
Obese	38.71 (0.75)	33.90*** (0.74)	35.00*** (0.08)	34.86*** (0.08)	32.71*** (0.09)	38.64 (0.09)

BMI = body mass index.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors are used in all calculations. For adults, a BMI below 18.5 is classified as underweight, between 18.5 and 24.9 is classified as normal weight, between 25.0 and 29.9 is classified as overweight, and 30.0 and above is classified as obese. The NHANES-outlier method excludes height and weight values that fall below or above the minimum and maximum height and weight for adults reported in NHANES. The IQR-outlier method excludes height and weight values that fall below or above the interquartile range (IQR). The predicted percentile rank method uses predictions of BMI based on a linear regression of measured BMI on percentile rankings of self-reported BMI in NHANES. Asterisk (*), double asterisk (**), and triple asterisk (***) indicate that the t-test of a difference compared to NHANES (measured) is significant at the 10-, 5-, and 1-percent levels, respectively.

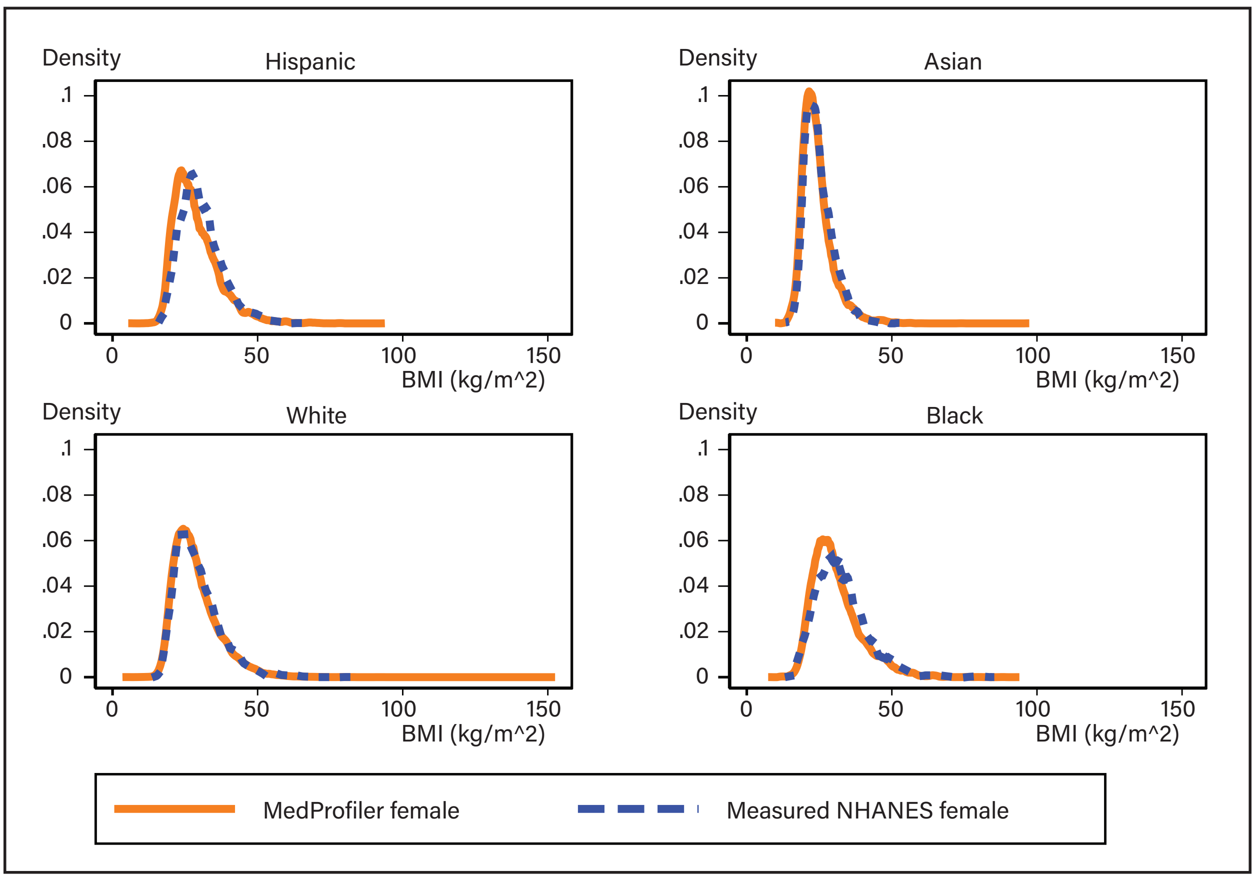
Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey.

Like many previous studies using reported measures, NHANES reported height is overreported and weight is underreported compared to measured height and weight values in NHANES. For 2011–2018 NHANES, the average reported height for adults 20 years and older is 66.9 inches (1.70 meters), a statistically significant 0.8 inches more than measured height (66.1 inches; 1.68 meters). Although average reported weight is 1.7 pounds (0.78 kilograms) less than its measured counterpart in NHANES, this difference is not statistically significant at the 5-percent level. Taken together, average BMI calculated using measured height and weight is 0.74 kg/m^2 more than BMI based on self-reported height and weight, and the population characterized as being obese is 5 percentage points less. Unlike the reported values in NHANES, both average reported height and weight in the MedProfiler adult sample are greater than measured height and weight in NHANES. But similar to average BMI based on reported height and weight in NHANES, BMI based on MedProfiler is less than average BMI based on measured height and weight in NHANES (28.9 versus 29.2 kg/m^2). Thus, less of the MedProfiler adult sample is classified as obese compared to NHANES (35 percent versus 39 percent).

Figure 3 compares the densities of adult BMI by dataset (measured NHANES and reported MedProfiler), race or ethnicity, and gender, and illustrates some differences between the distributions. First, reported BMI in the MedProfiler tends to have larger standard deviations and be more skewed than the measured BMI for most of the gender and demographic profiles (table A.4). In addition, the median and 90th percentiles for reported BMI in the MedProfiler are generally less than those using measured BMI in NHANES (table 4). Last, the BMI distribution based on the MedProfiler has more extreme BMI values (figure 3) and larger kurtoses compared to measured BMI in NHANES (table A.4). Consistent with the comparison of sample moments, the KS test indicates that the distribution of reported BMIs for adults in the MedProfiler is not drawn from the same population distribution as the measured BMIs in NHANES (table 4).

Figure 3a

Measured (NHANES) and reported (MedProfiler) BMI (kg/m^2) for adult females by demographic group



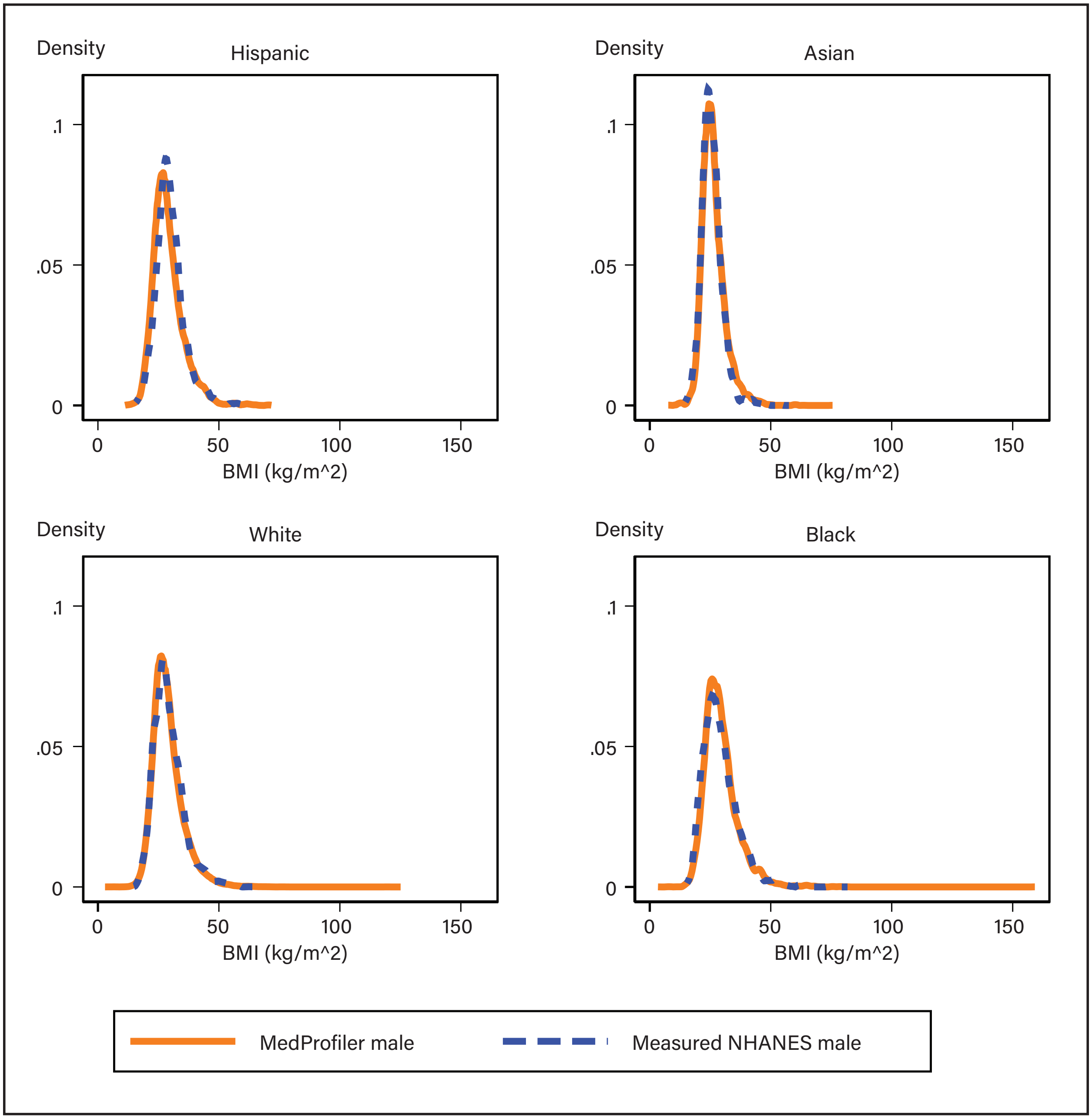
BMI = body mass index.

Note: Sample weights and projection factors are used in all calculations.

Source: USDA, Economic Research Service calculations based on Epanechnikov kernel density estimations (half-width of kernel equal to 3) and the 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey data.

Figure 3b

Measured (NHANES) and reported (MedProfiler) BMI (kg/m²) for adult males by demographic group



BMI = body mass index.

Note: Sample weights and projection factors are used in all calculations.

Source: USDA, Economic Research Service calculations based on Epanechnikov kernel density estimations (half-width of kernel equal to 3) and the 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey data.

Table 4

Mean, percentiles, and tests of distributions of measured (NHANES) and reported (MedProfiler) adult BMI (kg/m^2) by gender and demographic group

	Mean		50th percentile		90th percentile		
	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)	KS test (D-statistic)
Male							
Hispanic	29.8	29.0	29.1	28.0	36.9	37.0	0.10***
White	29.1	28.9	28.2	27.9	36.7	36.9	0.03***
Black	29.0	29.4	28.0	28.2	38.1	38.0	0.06***
Asian	25.7	26.3	25.2	25.7	30.9	32.3	0.08***
Other race or ethnicity	30.0	28.7	28.8	27.5	39.5	36.8	0.09***
Female							
Hispanic	30.4	28.8	29.2	27.3	40.0	39.0	0.16***
White	29.0	28.8	27.6	27.3	39.3	39.1	0.04***
Black	32.3	30.8	31.2	29.2	43.4	42.1	0.12***
Asian	24.7	24.3	23.8	23.2	31.3	30.9	0.06***
Other race or ethnicity	30.3	29.6	28.8	27.5	41.7	41.2	0.10***

BMI =body mass index.

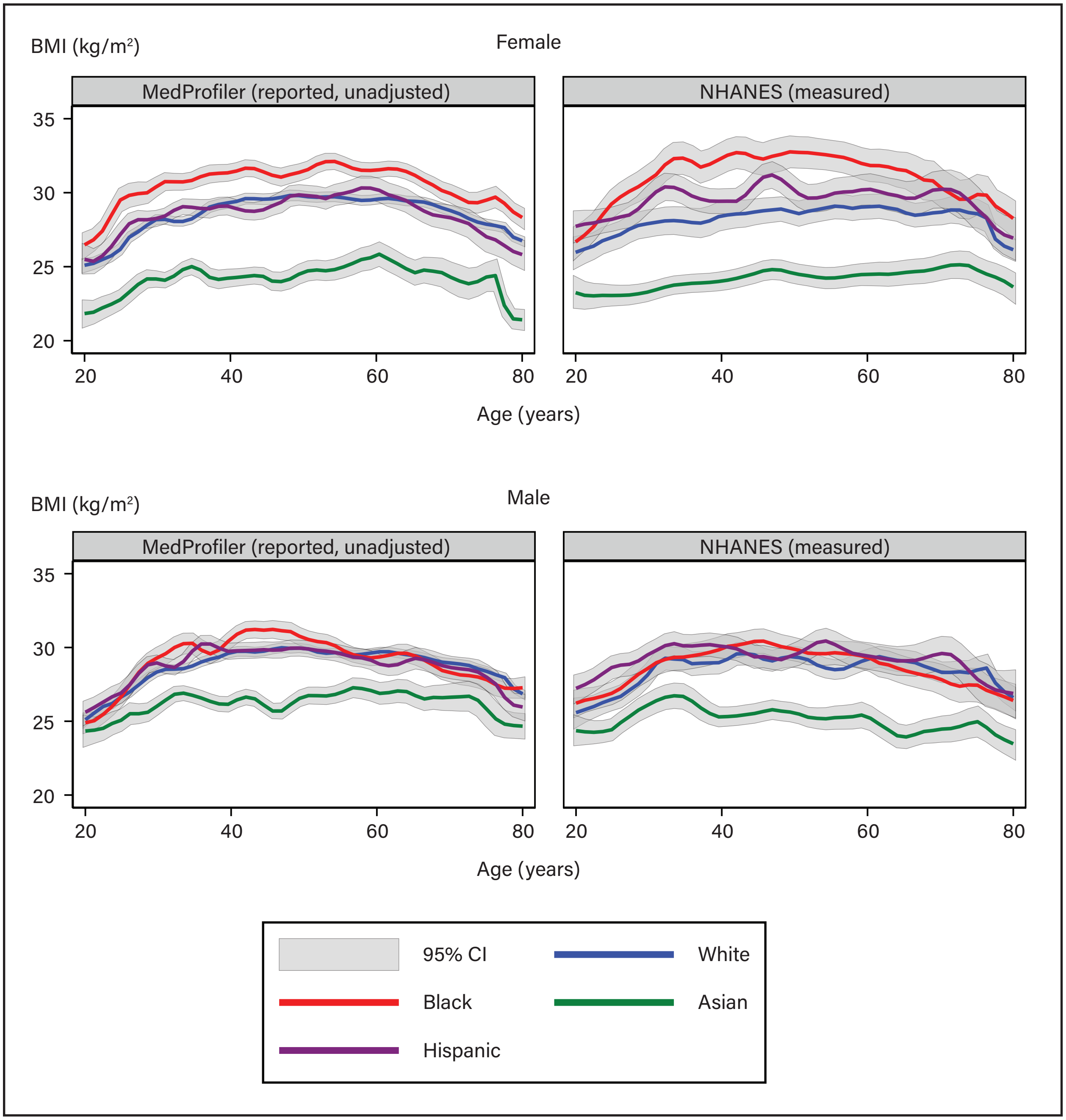
Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors are used in all calculations. The Kolmogorov-Smirnov (KS) test compares the distribution of BMIs based on MedProfiler reported height and weight to the NHANES measured values. Asterisk (*), double asterisk (**), and triple asterisk (***) indicate that the D-statistic is significant at the 10-, 5-, and 1-percent levels, respectively.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey data.

Figure 4 compares the relationship between age and BMI between the two datasets (measured in NHANES and reported in MedProfiler). Both datasets show the expected pattern between age and BMI: an inverse U-shaped relationship across all demographic groups, with Asian males and females having the lowest BMI across all ages. However, measured NHANES shows that White females have BMIs lower than Hispanic and Black females for most ages (especially under 60), but in the MedProfiler the BMIs for these groups are roughly the same.

Figure 4

Relationship between age and mean BMI (kg/m²) for adults in NHANES (measured) and MedProfiler (reported) by gender and demographic group



BMI = body mass index.

95% CI = 95-percent confidence interval. This measure indicates that there is 95 percent confidence that the true population parameter resides in this range.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors are used in all calculations.

Source: USDA, Economic Research Service calculations based on local polynomial smoothing and the 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–18 IRI MedProfiler Survey data.

Overall, misreporting of height and weight—hence BMI—appears less prevalent in the adult sample compared to the child and youth sample in the IRI MedProfiler. This is likely because projection factors in the IRI are constructed based on matching the household heads' demographic characteristics to U.S. Census demographic targets. The BMI distributions for White and Black male adults and White female adults in the

MedProfiler align well with the measured BMI in NHANES in terms of median and 90th percentiles and standard deviation (table 4 and table A.4). However, the BMI distributions for the other demographic profiles differ between the reported values in the MedProfiler and the measured values in NHANES, and kurtosis is larger for almost all demographic-gender groups as well. Finally, KS tests indicate that distributions of BMI for all gender and demographic groups based on the MedProfiler are not the same as their counterparts in NHANES (table 4).

Adjustments for Reducing Measurement Bias

To reduce measurement bias from reported rather than measured values in calculating BMI, the authors first explored truncating the full adult MedProfiler sample to exclude outliers based on NHANES and IQR cutoffs. Using the minimum and maximum measured height and weight by gender and demographic groups in NHANES as cutoffs, 2,282 adults were excluded. Using the IQR-outlier method, retaining only height and weight values within the IQR range, excluded 17,447 adults.

Both outlier methods reduce average BMI in the adult MedProfiler sample, and the proportion of adults categorized as obese, further exacerbating differences compared to measured NHANES values (column 3 of table 3). Compared to the unadjusted MedProfiler sample, the NHANES outlier method reduces average BMI by less than 0.1 kg/m^2 and the IQR outlier method reduces average BMI by roughly 0.7 kg/m^2. The relatively larger reduction in mean BMI based on the IQR outlier method compared to the NHANES outlier method resulted in a larger reduction of adults classified as obese.

The IQR outlier method also results in relatively larger reductions in points along (i.e., median and 90th percentile) and moments of (i.e., standard deviation, skewness, and kurtosis) the reported BMI distributions for all demographic groups and gender profiles (tables 4 and 5; tables A.4 and A.5). These relatively large changes move the BMI distributions away from the measured BMI distributions in NHANES. Hence, the IQR outlier method may over-adjust some demographic-gender BMI distributions, which exacerbates differences between BMI in NHANES and MedProfiler. The KS tests further show that the truncated reported BMI distributions based on both outlier methods are not from the same population as the measured BMI distributions based on NHANES (table 5).

Table 5

Mean, percentiles, and tests of distributions of reported adult BMI (kg/m^2) in MedProfiler by gender, demographic group, and adjustment method

	Mean			50th percentile			90th percentile			KS test (D-statistic)		
	Outlier methods		Predicted percentile rank method	Outlier methods		Predicted percentile rank method	Outlier methods		Predicted percentile rank method	Outlier methods		Predicted percentile rank method
	NHANES	IQR		NHANES	IQR		NHANES	IQR		NHANES	IQR	
Male												
Hispanic	28.90	28.30	29.11	27.89	27.60	28.18	36.91	35.43	36.51	0.18***	0.19***	0.02
White	29.01	28.33	29.74	27.97	27.70	29.18	36.95	35.42	36.51	0.28***	0.30***	0.02
Black	29.40	28.71	29.03	28.20	27.99	27.74	38.02	36.58	37.70	0.14***	0.15***	0.02
Asian	26.35	25.91	25.69	25.69	25.46	25.17	32.27	31.32	30.44	0.17***	0.18***	0.03
Other	28.64	27.83	29.88	27.47	27.26	28.62	36.61	34.45	39.60	0.23***	0.24***	0.04
Female												
Hispanic	28.72	28.08	29.03	27.29	26.78	27.61	39.05	37.76	39.26	0.15***	0.16***	0.02
White	28.56	27.98	30.36	27.25	26.63	29.22	38.40	37.03	39.67	0.28***	0.29***	0.02
Black	30.79	30.09	32.30	29.17	28.97	31.21	41.96	40.23	43.29	0.20***	0.22***	0.02
Asian	24.18	23.59	24.67	23.17	23.02	23.65	30.66	29.26	30.96	0.18***	0.17***	0.05***
Other	29.52	28.53	30.21	27.46	27.27	29.75	40.80	38.27	41.12	0.18***	0.20***	0.05

BMI = body mass index.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors are used in all calculations. The NHANES-outlier method excludes height and weight values that fall below or above the minimum and maximum height and weight for adults reported in NHANES. The IQR-outlier method excludes height and weight values that fall below or above the interquartile range (IQR). The predicted percentile rank method uses predictions of BMI based on a linear regression of measured BMI on percentile rankings of self-reported BMI in NHANES. The Kolmogorov-Smirnov (KS) test compares the distribution of adjusted BMIs to the distribution of measured BMIs in NHANES. Asterisk (*), double asterisk (**), and triple asterisk (***) indicate that the D-statistic is significant at the 10-, 5-, and 1-percent levels, respectively.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey.

Unlike the outlier methods, the distributions of predicted BMI using the percentile-ranking method are similar to the distributions of measured BMI from NHANES across demographic and gender profiles (table 5). (For percentile-rank regression coefficients, see table A.6). The average predicted BMI using the percentile rank method (29.3 kg/m^2) is not statistically different from its measured counterpart in NHANES (29.2 kg/m^2) (table 3). Similarly, the distribution of adults categorized as normal weight, overweight, and obese using measured NHANES BMI compared to distributions using predicted percentile rank BMI are not statistically different (table 3).[18] With the exception of White male and female adults, the average and median predicted BMI for the different demographic-gender groups are closer in magnitude to their measured counterparts compared to the unadjusted and outlier-adjusted MedProfiler samples (tables 4 and 5).

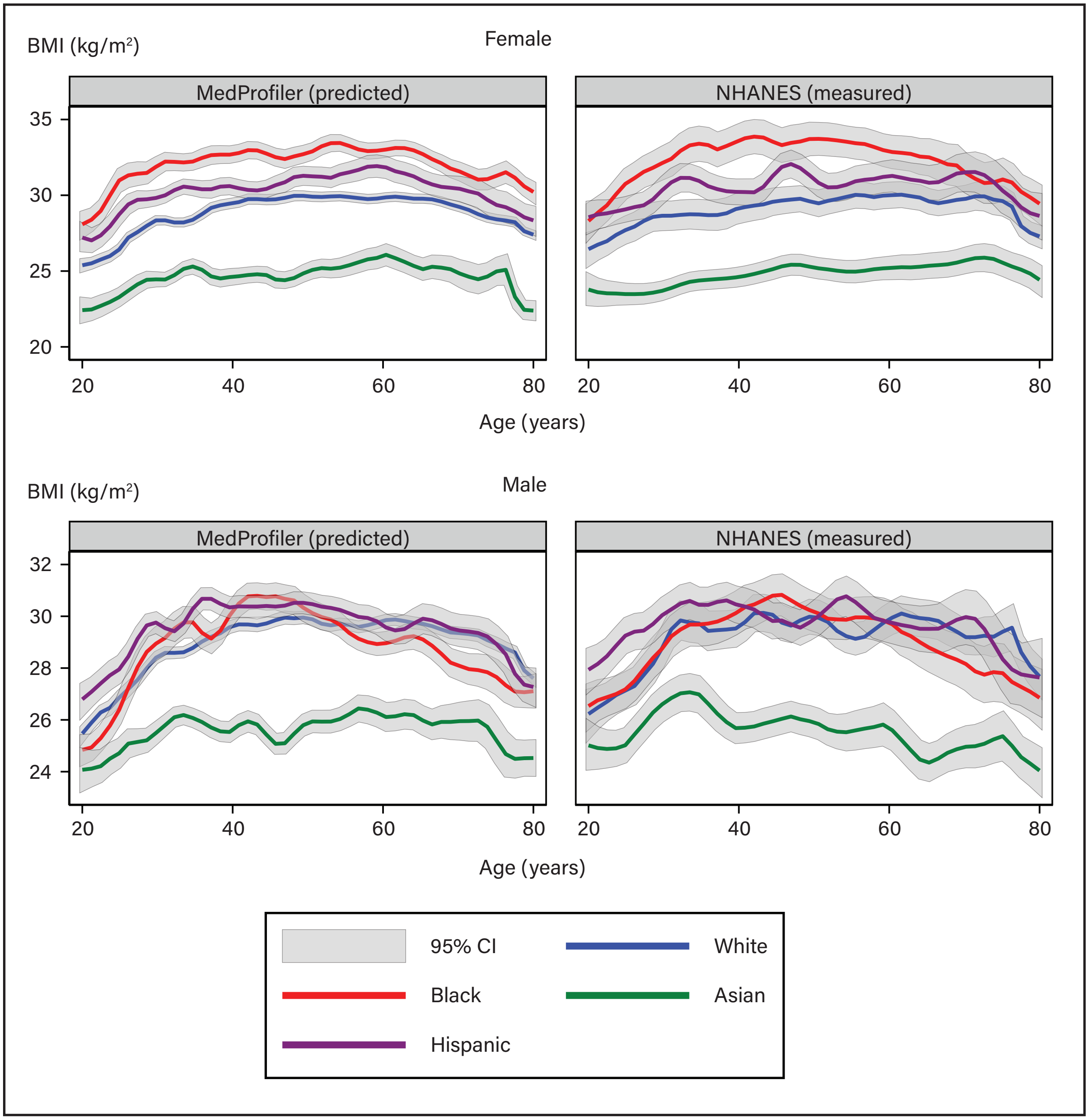
Like the NHANES-outlier method, the standard deviation, skewness, and kurtosis for predicted BMI based on the MedProfiler are in line with measured BMI upper moments from NHANES (tables A.4 and A.5). However, unlike both outlier methods, the KS tests indicate that the distribution of predicted BMI in the MedProfiler based on the percentile-ranking regression for almost all gender and demographic groups is the same as the distribution of measured BMI in NHANES (table 5). The one exception is Asian women; however, even for this group, the predicted percentile rank method still aligns best along all points (median and 90th percentile) and moments of the distribution with the measured BMI distributions from NHANES.

To better understand the impact of the percentile-ranking adjustments, the authors compared the relationship between age and BMI based on measured BMI from NHANES and predicted BMI using reported values from the MedProfiler (figure 5). Overall, the relationships observed in NHANES are maintained using the percentile-ranking predictions. For example, in NHANES, female BMI increased with age until approximately age 60 and then decreased for most demographic groups, with Asian females having the lowest average BMI and Black females having the highest. Although BMI in the unadjusted MedProfiler sample maintained some of these relationships, the predicted BMI more closely mimics the measured age-BMI relationships based on NHANES.

[18] The one exception is the share of adults categorized as underweight (1.1 percent), which is significantly lower than the share using measured NHANES BMI (1.5 percent).

Figure 5

Relationship between age and BMI (kg/m^2) for adults in NHANES (measured) and predicted MedProfiler (reported) by gender and demographic group



BMI = body mass index.

95% CI = 95-percent confidence interval. This measure indicates that there is 95 percent confidence that the true population parameter resides in this range.

Note: Sample weights and projection factors used in all calculations. The BMI reported in MedProfiler is adjusted using predictions from the linear regression of measured BMI (NHANES) on percentile rankings of self-reported BMI (NHANES).

Source: USDA, Economic Research Service calculations based on local polynomial smoothing and the 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012-2018 IRI MedProfiler Survey data.

Overall, BMI adjustments for the MedProfiler using percentile rankings of self-reported BMI in NHANES to predict measured BMI work well in practice. This is the only method for which the KS tests indicate that the distributions of predicted BMI are, in general, the same as the distributions of measured BMI in NHANES. While the outlier methods are simpler, the IQR-outlier method may overadjust the IRI MedProfiler sample, exacerbating differences along points in self-reported and measured BMI distributions. However, in some situations adjusting the sample by excluding outliers may be the only method available, as is the case for children.

Defining Household Obesity Status

A central challenge to studying the relationship between diet and obesity is that grocery purchases are often made at the household level and body mass is measured at the individual level. Thus, the question arises as to how to define obesity at the household level. Studies have answered this question in several ways (Doak et al., 2002; Staudigel, 2012; Chen et al., 2016; Jo, 2017; Volpe et al., 2019), and there is no clear consensus. To better understand the implications of different definitions on the share of households classified as obese, the researchers drew from existing literature and conducted a simple comparison. Using the percentile-ranking adjustment method for adults and no adjustments for children and youths, they compared household obesity for the MedProfiler sample using four distinct definitions: (1) the primary shopper is classified as obese, (2) any household member is classified as obese, (3) at least half of household members are classified as obese, and (4) all household members are classified as obese.

Given special interest in households with children and childhood obesity, the researchers considered two additional definitions for the subset of households with children: (1) any child or youth in the household is classified as obese, and (2) at least one adult and one child or youth in the household is classified as obese. Table 6 shows the share of all households classified as obese by household size and by race and ethnicity, as well as the share of all households with children or youths by race and ethnicity.

Table 6

Share of households classified as obese by household obesity definition, household size, and race and ethnicity (percent)

	(1) Primary shopper	(2) Any member	(3) 50 percent of members	(4) 100 percent of members
All households (n = 165,535)	41.59	55.96	47.16	22.76
Household size				
1	42.02	42.02	42.02	42.02
2	39.53	55.77	55.77	19.43
3	42.36	66.12	31.44	8.59
4	38.43	66.84	34.26	3.49
5	39.53	71.52	21.05	1.98
6	47.80	75.69	25.80	1.23
7	41.89	74.58	17.95	0.17
8+	46.77	83.57	24.24	0.33
Race and ethnicity				
Hispanic	47.35	67.21	45.17	16.79
White	38.47	53.84	41.92	20.78
Black	56.48	67.66	54.42	30.93
Asian	13.30	27.28	13.51	3.20
Other race or ethnicity	47.62	64.09	49.90	25.77
	(1) Primary shopper	**(2) Any child**	**(3) At least one adult and one child**	**(4) 100 percent of members**
All households with children or youths (n = 30,314)	40.69	26.63	18.76	5.13
Race and ethnicity				
Hispanic	48.63	32.19	23.39	4.64
White	36.52	27.69	18.46	4.29
Black	58.27	36.55	28.64	8.96
Asian	14.93	21.46	7.61	0.92
Other race or ethnicity	47.35	32.36	24.53	6.03

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Race and ethnicity of the household is based on the race and ethnicity of the primary shopper. Sample weights and projection factors are used in all calculations. BMIs for adult household members (age 20 and older) were adjusted using predicted-percentile-ranking method and with no adjustments for children and youths (aged 2 to 19). Household size is calculated based on the full household before dropping NHANES outliers. For adults, a BMI of 30.0 and above is classified as obese. For children and youths, body weight classification is based on age and gender (see Centers for Disease Control and Prevention, National Center for Health Statistics, 2021).

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey.

The proportion of households classified as obese differs substantially across definitions and household characteristics. Across all households, household obesity defined using the primary shopper, any member of the household, or at least half of household members, results in household obesity rates ranging between 42 and 56 percent. In contrast, by using the most conservative definition and defining household obesity using all members of the household, the percent of households classified as obese is 23 percent, reduced by more than half.

Definitions of household obesity are sensitive to household size. For single-person households, obesity rates are 42 percent across all definitions. However, as the number of household members increases, variation in household obesity rates also increases. In particular, household obesity rates increase with household size when defining household obesity by any household member. Conversely, household obesity rates decrease as household size increases when defining household obesity by percentage of household members—either 50 percent or 100 percent. For example, for a household of three, this can lead to very large differences, with 67 percent of households classified as obese using the threshold of any member compared to 31 percent using the 50 percent threshold. When considering the obesity status of the primary shopper only, there is no clear association with household size.

Household obesity rates also vary by definition when considering the race and ethnicity of the household, defined here as the race and ethnicity of the primary shopper. Using the primary shopper definition, approximately 38 to 56 percent of households across all race and ethnicity groups are classified as obese, except for Asian households (13 percent). The pattern is similar when defining household obesity by half of all household members, where 42 to 54 percent of all households are classified as obese and the percent of Asian households classified as obese is 14 percent. Defining household obesity by any household member results in approximately two-thirds of Hispanic, Black, and other racial and ethnic households classified as obese; this pattern does not hold for White and Asian households, 54 and 27 percent, respectively. Finally, defining household obesity by all household members results in a larger share (31 percent) of Black households classified as obese compared to other racial and ethnic households (26 percent), White households (21 percent), Hispanic households (17 percent), and Asian households (3 percent).

For households with children and youths, defining household obesity by the primary shopper results in the most households classified as obese across all racial and ethnic groups, except Asian households. Based on this definition, 41 percent of households with children and youths are classified as obese, whereas the other definitions result in markedly lower household obesity rates: 27 percent based on "any child," 19 percent based on "at least one child and one adult," and only 5 percent based on "all household members." By race and ethnicity, households follow a similar pattern for all household obesity definitions: Black households have the highest household obesity rates, followed by Hispanic and other racial and ethnic households.

Overall, household obesity rates in the full sample and the subset of households with children vary substantially across definitions. Data availability, research aims, and the implications of the chosen definition are important considerations when defining household obesity status. Alternatively, defining household obesity by the obesity status of any member increases the percentage of households classified as obese across all household sizes and all racial and ethnic groups. When focusing on households with children, the ordinal ranking of obesity rates across race and ethnicity is similar across all definitions, but the percentage classified as obese is much larger when defining household obesity by the primary shopper compared to other definitions that include children. Ideally, researchers may consider using multiple definitions to identify bias introduced into an analysis coming from a specific approach to defining household obesity.

Conclusion

The MedProfiler survey contains health information for individuals who participate in the IRI Consumer Network household panel, allowing researchers to link household food purchases to self-reported height and weight of household members. These data offer a promising tool to study links between food purchases and health outcomes. Before this study, little information was available about the quality of the MedProfiler self-reported data. This analysis compares BMI from self-reported height and weight in the MedProfiler data to their measured counterparts in NHANES to assess the value of MedProfiler data for food and health policy research.

Compared to measured BMI in NHANES, self-reported BMI based on the MedProfiler differs importantly across gender, racial and ethnic groups, and age. For some moments and points along the self-reported BMI distributions in the MedProfiler, some racial-ethnic and gender groups are consistent with measured BMI. For example, mean BMI, standard deviations, and 90th percentiles for White and Black male adults and White female adults in the MedProfiler align with their measured counterparts in NHANES. However, other racial and ethnic adult groups, as well as children and youth, show larger differences in lower moments and percentiles, and all subpopulations differ markedly at the upper moments of their BMI distribution.

The researchers explored several methods for adjusting BMI distributions to reduce measurement bias associated with self-reported data. For adults, predictions of measured BMI in the MedProfiler based on measured and self-reported percentile rankings in NHANES performed well. This percentile-ranking regression method developed by Courtemanche et al. (2015) is compatible with differential error as long as both samples are representative of the same population. KS tests indicate that the distribution of predicted BMI in the MedProfiler based on the percentile ranking regression for almost all gender and demographic groups and moments (standard deviation, skewness, and kurtosis) is the same as the distribution of measured BMI in NHANES. Although the KS test indicates that BMI for Asian females is an exception, moments of and movements along the predicted BMI distribution for Asian females are more consistent with measured BMI compared to unadjusted reported BMI. Because this method maintains the full sample and adjusts reported BMI in the MedProfiler to best align with measured BMI in NHANES, the authors recommend using the method when a validation dataset with both measured and reported values is available for the same population.

Other adjustment methods applied to the MedProfiler sample removed individuals with BMI outside of cutoffs based on the IQR of self-reported BMI values or minimum- and maximum-measured BMI values in NHANES. While these outlier methods are simpler, using an IQR-outlier method may overadjust the MedProfiler sample, exacerbating differences between self-reported and measured BMI distributions. However, in some instances, adjusting the sample by excluding outliers may be the only method available. This is the case for children and youth, but evidence from this study suggests neither outlier method is superior for correcting bias. Other methods could be employed to adjust BMI based on self-reported height and weight, such as imputing outlier values rather than excluding them. Future analysis warrants examining whether alternative imputation methods like top- and bottom-coding, hot-deck or cold-deck imputing, or mean substituting of outlier values offer better methods of dealing with measurement issues related to self-reported values.

Finally, since scanner data include food purchases by households rather than individuals, the researchers considered methods for aggregating a household's body weight status. By calculating the percentage of households identified as obese using four different definitions for both the full sample of households and a subsample of households with children, they find that choice of definition can strongly impact which households and how many are included. The researchers urge future investigators to consider carefully which definition fits the question at hand and the implications of the chosen definition. When feasible, more than one method should be used and presented to ensure transparency in reporting possible biases. Further work is needed to better understand biases in outcomes using varied household body weight definitions.

There are a few limitations to the analyses in this report. First, several years of NHANES and MedProfiler data were stacked together to achieve a sample size large enough to examine some race/ethnicity and gender groups. Use of multi-year samples requires an assumption that BMI did not change substantially from 2012 to 2018. This assumption may not be completely supported since evidence suggests it holds for latter years in the study but not earlier years (Hales et al., 2017). When the researchers tested the implications of this limitation in a sensitivity analysis for adults limited to years 2017 and 2018 (table A.7), they found qualitatively similar results—the predicted percentile-ranking method performed better than both outlier methods. Second, this analysis does not show how corrections for misreporting of BMI affect empirical research. For example, the preferred method for adults, predicted BMI using the percentile-ranking regression method, increases the proportion of the MedProfiler adult sample classified as obese by almost 5 percentage points. It remains to be seen how this reclassification of adults from a nonobese to an obese category affects modeling results. Forthcoming USDA, ERS research will analyze associations between unadjusted and adjusted body weight status and purchasing behavior.

Last, the MedProfiler projection factors are constructed to match demographic characteristics of the household's primary shopper, which assumes that households are homogenous in terms of demographic makeup. Although the multiracial U.S. household population is still relatively small (about 10 percent in 2020), it more than tripled from approximately 3 percent in 2010 (Jones et al., 2021). This implies that household composition is becoming less homogenous. Hence, researchers should be careful when interpreting results concerning race and ethnicity and body weight status of respondents who are not primary shoppers.

While the difference between BMI and obesity estimates in NHANES compared with the MedProfiler and across differing definitions of household obesity is substantial, the impact on economic research on consumption remains to be seen. Forthcoming work focuses on food purchasing patterns by the BMI status of a household, including subanalyses for areas of economic importance such as income level. Sensitivity analyses using unadjusted data and differing definitions of household obesity will offer an opportunity to understand the quantitative value of BMI adjustments in scanner data as well as decisions related to household body weight classification.

References

Bolton-Smith, C., M. Woodward, H. Tunstall-Pedoe, and C. Morrison. 2000. "Accuracy of the Estimated Prevalence of Obesity From Self-Reported Height and Weight in an Adult Scottish Population," *Journal of Epidemiology Community Health* 54:143–48.

Carlson, A., C.E. Tornow, E.T. Page, A.B. McFadden, and T.P. Zimmerman. 2022. "Development of the Purchase to Plate Crosswalk and Price Tool: Estimating Prices for the National Health and Nutrition Examination Survey (NHANES) Foods and Measuring the Healthfulness of Retail Food Purchases," *Journal of Food Composition and Analysis* 106:104344.

Carlson, A., E.T. Page, T.P. Zimmerman, C.E. Tornow, and S. Hermansen. 2019. *Linking USDA Nutrition Databases to IRI Household-Based and Store-Based Scanner Data*, Technical Bulletin 1952, U.S. Department of Agriculture, Economic Research Service, March 2019.

Centers for Disease Control and Prevention, National Center for Health Statistics. 2017. *Clinical Growth Charts,* available online.

Centers for Disease Control and Prevention, National Center for Health Statistics. 2021. *Defining Childhood Weight Status,* available online.

Centers for Disease Control and Prevention, National Center for Health Statistics. *NHANES 2017–2018 Examination Data Overview,* available online.

Chen, D., E.C. Jaenicke, and R.J. Volpe. 2019. "The Healthfulness of Food-at-Home Expenditures, the Local Food Environment, and Childhood Obesity," *American Journal of Health Promotion* 33(3):412–19.

Chen, T.C., J.D. Parker, J. Clark, H.C. Shin, J.R. Rammon, and V.L. Burt. 2018. "National Health and Nutrition Examination Survey: Estimation Procedures, 2011–2014," *National Center for Health Statistics Vital Health Statistics* 2(177).

Courtemanche, C., J.C. Pinkston, and J. Stewart. 2015. "Adjusting Body Mass for Measurement Error With Invalid Validation Data," *Economics and Human Biology* 19:275–93.

Doak, C., L. Adair, M. Bentley, Z. Fengying, and B. Popkin. 2002. "The Underweight/Overweight Household: An Exploration of Household Sociodemographic and Dietary Factors in China," *Public Health Nutrition* 5(1A):215–21.

Dong, D., H. Stewart, and P. McLaughlin. 2018. "A New Approach to Estimate Household Food Demand With Panel Data," *Journal of Applied Economics* 21(1):122–36.

Ezzati, M., H. Martin, S. Skjold, S. Vander Hoorn, and C.J. Murray. 2006. Trends in National and State-Level Obesity in the USA After Correction for Self-Report Bias: Analysis of Health Surveys. *Journal of the Royal Society of Medicine*, *99*(5):250–57.

Flegal, K.M., C.L. Ogden, C. Fryar, J. Afful, R. Klein, and D.T. Huang. 2019. "Comparisons of Self-Reported and Measured Height and Weight, BMI, and Obesity Prevalence From National Surveys: 1999–2016," *Obesity* 27(10):1711–19.

Gorber, S. Connor, M. Tremblay, D. Moher, and B. Gorber. 2007. "A Comparison of Direct Vs. Self-Report Measures for Assessing Height, Weight, and Body Mass Index: A Systematic Review," *Obesity Reviews* 8(4):307–26.

Hales, C.M., M.D. Carroll, C.D. Fryar, and C.L. Ogden. 2017. *Prevalence of Obesity Among Adults and Youth: United States, 2015–16*, National Center for Health Statistics Data Brief 288.

Jo, Y. 2017. *The Differences in Characteristics Among Households With and Without Obese Children: Findings From USDA's FoodAPS*, EIB-179, U.S. Department of Agriculture, Economic Research Service, September 2017.

Jones, N., R. Marks, R. Ramirez, and M. Rios-Vargas. 2021. *2020 Census Illuminates Racial and Ethnic Composition of the Country*, U.S. Department of Commerce, Bureau of the Census, available online.

Keith, S., K. Fontaine, N. Pajewski, T. Mehta, and D. B. Allison, 2011. "Use of Self-Reported Height and Weight Biases the Body Mass Index-Mortality Association," *International Journal of Obesity* 35(3):401–08.

Kuczmarski, M.F., R.J. Kuczmarski, and M. Najjar. 2001. "Effects of Age on Validity of Self-Reported Height, Weight, and Body Mass Index: Findings From the Third National Health and Nutrition Examination Survey, 1988–1994," *Journal of the American Dietetic Association* 101(1):28–36, available online.

Kuhns, A., and M. Saksena. 2017. *Food Purchase Decisions of Millennial Households Compared to Other Generations*, EIB-186, U.S. Department of Agriculture, Economic Research Service, December 2017.

Lyons, A., J. Park, and C. H. Nelson, 2008. "Food Insecurity and Obesity: A Comparison of Self-Reported and Measure Height and Weight," *American Journal of Public Health* 98(4):751–57.

Mancino, L., J. Guthrie, M. Ver Ploeg, and B.-H. Lin. 2018. *Nutritional Quality of Foods Acquired by Americans: Findings From USDA's National Household Food Acquisition and Purchase Survey*, EIB-188, U.S. Department of Agriculture, Economic Research Service, February 2018.

Massey, F.J. 1951. "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association* 46(253):68–78.

Muth, M.K., M. Sweitzer, D. Brown, K. Capogrossi, S. Karns, D. Levin, A. Okrent, P. Siegel, and C. Zhen. 2016. *Understanding IRI Household-Based and Store-Based Scanner Data*. Technical Bulletin 1942, U.S. Department of Agriculture, Economic Research Service, April 2016.

Ogden, C.L., C.D. Fryar, C.B. Martin, D.S. Freedman, M.D. Carroll, Q. Gu, and C.M. Hales. 2020. "Trends in Obesity Prevalence by Race and Hispanic Origin—1999–2000 to 2017–2018," *Journal of the American Medical Association (JAMA)* 324(12):1208–10.

Rahkovsky, I., and S. Snyder. 2015. *Food Choices and Store Proximity*, ERR-195, U.S. Department of Agriculture, Economic Research Service, September 2015.

Rowland, M.L. 1990. “Self-Reported Weight and Height,” *The American Journal of Clinical Nutrition* 52(6):1125–33.

Spencer, E.A., P.N. Appleby, G. K. Davey, and T.J. Key. 2002. “Validity of Self-Reported Height and Weight in 4,808 EPIC-Oxford Participants,” *Public Health Nutrition* 5(4):561–65.

Staudigel, M. 2012. “How Do Obese People Afford to be Obese? Consumption Strategies of Russian Households,” *Agricultural Economics* 43:701–14.

Stommel, M., and C.A. Schoenborn. 2009. “Accuracy and Usefulness of BMI Measures Based on Self-Reported Weight and Height: Findings From the NHANES & NHIS 2001–2006,” *BMC Public Health* 9(1):421.

Sweitzer, M., D. Brown, S. Karns, M.K. Muth, P. Siegel, and C. Zhen. 2017. *Food-at-Home Expenditures: Comparing Commercial Household Scanner Data From IRI and Government Survey Data*. Technical Bulletin 1946, U.S. Department of Agriculture, Economic Research Service, September 2017.

U.S. Burden of Disease Collaborators. 2018. “The State of U.S. Health, 1990–2016: Burden of Diseases, Injuries, and Risk Factors Among U.S. States,” *Journal of the American Medical Association (JAMA)* 319(14):1444–72.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. 2020. *Executive Summary of the Dietary Guidelines for Americans, 2020–2025*, 9th Edition, available online.

Volpe, R., A. Kuhns, and T. Jaenicke. 2017. *Store Formats and Patterns in Household Grocery Purchases*, EIB-167, U.S. Department of Agriculture, Economic Research Service, March 2017.

Volpe, R., and A. Okrent. 2012. *Assessing the Healthfulness of Consumers’ Grocery Purchases*, EIB-102, U.S. Department of Agriculture, Economic Research Service, November 2012.

Volpe, R., X. Cai, C. Schroeter, and L. Mancino. 2019. “The Effect of Produce Purchasing Behavior on Adult Obesity and Body Weight: Evidence From Household-Level Scanner Data,” *Agricultural and Resource Economic Review* 48(2):253–73.

Ward, Z.J., S.N. Bleich, A.L. Cradock, J.L. Barrett, C.M. Giles, C. Flax, M.W. Long, and S.L. Gortmaker. 2019. “Projected U.S. State-Level Prevalence of Adult Obesity and Severe Obesity,” *The New England Journal of Medicine* 382(25):2440–50.

Waters, H., and M. Graf. 2018. *America’s Obesity Crisis: The Health and Economic Costs of Excess Weight,* Milken Institute.

Wen, M., and L. Kowaleski-Jones. 2012. “Sex and Ethnic Differences in Validity of Self-Reported Adult Height, Weight and Body Mass Index,” *Ethnicity and Disease* 22(1):72–78.

Wirfält, E., I. Drake, and P. Wallström. 2013. “What Do Review Papers Conclude About Food and Dietary Patterns?” *Food and Nutrition Research* 57(1).

World Health Organization (WHO). 2018. *Noncommunicable Diseases Country Profiles 2018.*

Wright, D.R., K. Glanz, T. Colburn, S.M. Robson, and B.E. Saelens. 2018. “The Accuracy of Parent-Reported Height and Weight for 6-12 Year Old U.S. Children,” *BMC Pediatrics* 18(52).

Zhen, C., E.A. Finkelstein, J. Nonnemaker, S. Karns, and J.E. Todd. 2014. “Predicting the Effects of Sugar-Sweetened Beverage Taxes on Food and Beverage Demand in a Large Demand System,” *American Journal of Agricultural Economics* 96(1):1–25.

Appendix

Table A.1

Upper moments of measured (NHANES) and reported (MedProfiler) BMI (kg/m^2) for children and youths by age and sex

Age and sex	Standard deviation		Skewness		Kurtosis	
	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)
Male						
2	1.5	7.5	1.2	3.6	7.2	26.4
3	2.1	6.1	3.1	5.7	18.8	59.9
4	1.8	5.0	1.8	3.6	8.2	24.8
5	2.1	5.7	1.6	7.6	6.5	147.4
6	2.7	5.2	2.8	3.0	16.3	19.9
7	3.2	4.5	1.8	2.2	7.9	13.8
8	3.6	4.9	1.5	1.8	6.1	10.3
9	3.9	4.7	1.5	1.7	5.7	10.9
10	4.6	4.8	1.4	1.5	5.9	7.1
11	4.9	11.0	1.0	28.3	4.1	961.4
12	4.7	4.9	1.2	1.2	4.3	5.7
13	5.6	4.8	1.2	1.2	4.3	5.4
14	5.4	14.1	1.3	41.4	5.1	1,940.1
15	6.1	5.0	1.2	1.3	4.6	5.7
16	5.8	5.4	1.7	1.5	6.6	6.7
17	6.2	5.5	1.5	1.4	5.6	6.2
18	7.1	5.5	1.7	1.5	6.5	6.4
19	5.7	5.8	1.0	1.2	3.6	4.7
Female						
2	1.4	7.4	0.8	3.9	4.8	31.8
3	1.6	6.4	3.5	7.3	32.3	103.6
4	2.0	4.7	1.3	3.3	5.4	26.5
5	2.1	5.2	1.6	4.2	6.1	39.0
6	3.1	5.1	1.6	2.7	6.3	14.8
7	3.3	4.6	1.4	1.8	5.4	9.0
8	3.8	5.3	1.8	2.5	8.9	14.6
9	4.2	5.0	1.3	2.9	4.8	26.7
10	4.3	4.8	1.1	1.7	5.2	9.1
11	5.1	5.9	1.1	3.6	4.0	30.7
12	5.5	4.6	1.0	1.3	4.1	6.3
13	5.4	5.0	1.4	1.6	5.3	7.5
14	5.2	5.6	1.2	2.4	4.6	19.3
15	5.8	5.7	1.4	1.8	5.3	9.1
16	6.5	5.6	1.4	1.5	5.0	6.0
17	6.3	5.8	1.3	1.5	5.0	5.8
18	7.7	6.1	1.9	1.5	8.4	5.6
19	7.5	6.5	1.4	1.4	5.7	5.1

BMI = body mass index.

Note: MedProfiler body weight data are reported by parents for children and youths under age 18 and self-reported by youths ages 18 and 19. Sample weights and projection factors are used in all calculations.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey.

Table A.2

Mean, percentiles, and upper moments of BMI (kg/m²) distributions for children and youths in the MedProfiler by sex, age, and adjustment method

Age and Sex	Mean		Percentiles: 50th		Percentiles: 90th		Standard deviation		Skewness		Kurtosis		Number of observations excluded	
	NHANES outlier	IQR outlier	NHANES outlier	IQR outlier	NHANES outlier	IQR outlier	NHANES outlier	IQR outlier	NHANES outlier	IQR outlier	NHANES outlier	IQR outlier	NHANES outlier	IQR outlier
Male														
2	16.9	18.8	16.6	17.1	20.6	26.9	2.6	5.9	0.7	1.8	4.2	6.7	391	47
3	16.6	16.5	16.3	16.3	19.5	20.1	3.3	3.1	2.1	1.1	13.5	6.4	183	108
4	16.1	16.3	15.9	15.9	19.9	20.2	2.8	3.1	0.9	0.8	5.8	4.8	279	97
5	16.0	16.3	15.8	15.9	19.3	20.8	3.0	3.5	1.5	1.3	8.6	6.8	248	90
6	16.6	16.2	15.6	15.5	21.2	20.5	3.8	3.3	1.9	1.2	9.0	6.0	187	149
7	16.4	16.2	15.6	15.6	20.8	20.4	3.5	3.2	1.5	0.9	7.5	4.6	238	218
8	17.6	17.3	16.5	16.5	23.0	22.7	4.3	3.9	1.4	1.0	6.2	4.5	129	75
9	18.2	17.9	17.4	17.2	23.4	23.2	4.3	3.9	1.3	0.8	6.4	4.0	68	64
10	19.1	18.6	18.2	17.9	24.7	23.6	4.6	3.9	1.5	0.9	6.5	4.4	86	99
11	19.5	19.0	18.8	18.7	24.9	24.0	4.3	3.7	1.1	0.7	5.0	3.9	56	137
12	20.3	20.0	19.5	19.5	26.4	25.7	4.4	4.2	1.0	0.8	4.3	3.8	149	118
13	21.2	20.7	20.2	19.9	27.4	26.6	4.6	4.3	1.2	0.8	5.0	3.6	94	88
14	21.7	21.3	20.6	20.4	28.3	27.4	4.7	4.2	1.1	0.7	4.3	3.3	110	100
15	22.4	21.7	21.5	21.2	29.0	27.8	4.9	4.1	1.3	0.7	5.4	3.5	65	111
16	22.9	22.3	21.9	21.7	29.8	28.1	5.3	4.2	1.5	0.7	6.5	3.5	65	105
17	23.5	22.8	22.4	22.1	30.6	28.6	5.2	4.2	1.4	0.9	5.5	4.1	53	126
18	23.9	23.3	23.0	22.7	30.7	29.4	5.2	4.3	1.4	0.8	5.9	3.8	56	127
19	24.5	24.2	23.3	23.1	32.5	31.6	5.4	5.0	1.0	0.9	3.9	3.8		
Female														
2	16.6	18.4	16.3	16.7	19.7	25.9	2.7	5.5	1.1	1.6	5.8	6.1	368	64
3	16.5	16.6	16.2	16.3	20.1	20.6	2.9	3.1	1.3	1.1	7.3	5.8	225	143
4	16.2	16.2	15.9	15.9	19.5	20.4	2.7	3.2	1.0	0.9	6.0	5.4	412	110
5	16.5	16.1	15.9	15.6	21.7	20.7	3.8	3.5	1.3	1.1	5.7	4.9	77	110
6	16.2	16.3	15.6	15.6	19.9	20.6	3.4	3.8	1.9	1.7	10.3	8.1	263	89
7	16.5	16.4	15.9	15.9	21.4	20.9	3.6	3.3	1.4	0.8	6.4	4.1	233	171
8	17.3	17.1	16.3	16.3	22.5	22.2	4.2	3.8	1.6	0.9	7.8	4.4	105	73
9	18.0	17.9	17.4	17.4	23.2	23.2	4.1	4.0	1.2	1.0	5.8	4.8	102	68
10	18.8	18.7	18.1	18.1	24.1	23.9	4.2	4.0	1.1	0.8	5.0	4.2	176	52
11	20.1	19.5	19.0	18.8	26.6	25.0	4.7	4.1	1.3	0.7	5.7	3.9	71	110
12	20.6	20.1	19.8	19.5	26.4	25.1	4.4	3.8	1.4	0.8	6.7	4.4	62	127
13	21.4	20.6	20.5	20.1	27.5	25.6	4.8	3.6	1.5	0.7	6.6	3.6	46	144
14	22.3	21.8	21.1	21.0	29.3	28.4	4.8	4.3	1.1	0.8	4.1	3.4	68	118
15	23.1	22.4	21.6	21.5	30.4	28.4	5.4	4.4	1.5	1.0	5.9	3.8	78	125
16	23.4	22.6	21.9	21.7	31.3	28.5	5.4	4.3	1.5	1.0	5.5	4.1	71	142
17	23.9	23.1	22.3	22.1	31.7	29.5	5.7	4.5	1.5	1.1	5.8	4.1	41	110
18	24.4	23.5	22.7	22.3	32.6	30.4	6.0	4.6	1.5	0.9	5.7	3.4	33	116
19	24.7	23.8	22.7	22.5	33.5	31.2	6.4	5.0	1.5	1.0	5.3	3.6	28	90

BMI = body mass index.

Note: MedProfiler body weight data are reported by parents for children and youths under age 18 and self-reported by youths ages 18 and 19. Sample weights and projection factors are used in all calculations.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey.

Table A.3

Obesity prevalence for children and youth by age and sex (percent)

Age and sex	NHANES (measured)	Unadjusted MedProfiler (reported)	Adjusted MedProfiler (reported)	
			NHANES outlier	IQR outlier
Male				
2	6.0	35.9	20.2	34.5
3	8.5	28.8	22.6	24.5
4	13.8	26.1	18.4	22.5
5	13.9	22.2	21.3	18.3
6	17.1	19.2	13.9	15.2
7	17.2	17.9	13.0	11.6
8	17.0	18.0	14.6	13.9
9	18.6	15.8	13.0	13.1
10	18.4	14.6	12.9	12.1
11	18.5	15.9	15.3	11.5
12	26.4	12.4	11.4	8.0
13	19.3	12.1	10.8	6.0
14	20.6	14.7	14.0	11.5
15	19.0	14.5	13.6	9.9
16	24.0	17.6	16.8	12.2
17	27.3	17.4	16.5	12.5
18	21.6	18.3	17.5	12.8
19	24.5	18.2	17.4	13.1
Female				
2	8.2	35.5	21.9	35.0
3	12.1	31.6	26.5	27.5
4	13.2	29.6	22.8	26.8
5	12.9	26.2	17.6	22.5
6	14.2	24.0	21.1	18.6
7	17.9	21.4	15.3	13.8
8	18.8	22.8	19.4	18.3
9	18.6	19.3	18.2	16.8
10	21.6	16.9	16.5	13.5
11	27.2	15.1	14.2	11.1
12	18.9	16.6	15.3	14.0
13	24.7	15.7	16.0	14.0
14	22.6	16.0	16.1	13.6
15	22.1	15.2	15.2	11.6
16	19.0	15.2	14.5	10.9
17	19.2	14.5	14.2	10.0
18	20.8	13.9	13.4	10.0
19	19.5	16.3	15.0	13.0

Note: MedProfiler body weight data are reported by parents for children and youths under age 18 and self-reported by youths ages 18 and 19. Sample weights and projection factors are used in all calculations. The NHANES-outlier method excludes height and weight values that fall below or above the minimum and maximum height and weight for children and youth by age and gender reported in NHANES. The IQR-outlier method excludes height and weight values that fall below or above the interquartile range (IQR) for children by age and gender.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey data.

Table A.4

Upper moments of measured (NHANES) and reported (MedProfiler) adult BMI (kg/m^2) distributions by demographic group

	Standard deviation		Skewness		Kurtosis	
	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)	NHANES (measured)	MedProfiler (reported)
Male						
Hispanic	5.7	6.3	1.0	1.3	5.2	6.7
White	6.1	6.3	1.5	1.4	10.0	8.0
Black	7.1	7.0	1.3	1.8	6.9	14.1
Asian	4.3	4.9	1.1	1.5	5.1	8.5
Other race or ethnicity	7.2	6.5	1.3	1.5	6.6	9.1
Female						
Hispanic	7.3	7.7	0.9	1.4	4.0	6.1
White	7.5	7.7	1.1	1.9	4.7	11.2
Black	8.6	8.3	1.0	1.2	4.9	5.1
Asian	4.8	5.3	1.2	1.3	5.0	6.2
Other race or ethnicity	8.3	8.5	1.4	1.4	6.8	5.9

BMI = body mass index.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors are used in all calculations.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey data.

Table A.5

Upper moments of the reported adult BMI (kg/m^2) distributions in the MedProfiler, by gender, race and ethnicity, and adjustment method

	Standard deviation			Skewness			Kurtosis		
	Outlier methods		Percentile ranking regression method	Outlier methods		Percentile ranking regression method	Outlier methods		Percentile ranking regression method
	NHANES	IQR		NHANES	IQR		NHANES	IQR	
Male									
Hispanic	6.1	5.1	5.3	1.2	0.6	1.1	5.5	3.4	4.7
White	6.1	5.2	5.9	1.2	0.6	0.9	5.6	3.4	4.6
Black	6.8	5.7	6.7	1.4	0.6	1.3	6.8	3.4	5.5
Asian	4.7	4	4	1.2	0.7	1.2	5.5	3.7	6.7
Other	6.3	5.1	6.9	1.3	0.5	1.1	5.9	3.4	5.5
Female									
Hispanic	7.3	6.4	6.9	1.2	0.8	1.1	5.1	3.3	4.7
White	7.6	6.6	7.3	1.1	0.7	0.8	4.3	3.2	3.7
Black	8.1	7	8.2	1.1	0.7	0.9	4.7	3.2	4.4
Asian	5.1	4.1	4.5	1.5	0.7	1.1	6.5	3.5	4.4
Other	8.3	6.9	7.8	1.4	0.8	1.1	5.6	3.2	4.8

BMI = body mass index.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors are used in all calculations. The NHANES-outlier method excludes height and weight values that fall below or above the minimum and maximum height and weight for adults reported in NHANES. The IQR-outlier method excludes height and weight values that fall below or above the interquartile range (IQR). The predicted percentile rank method uses predictions of BMI based on a linear regression of measured BMI on percentile rankings of self-reported BMI in NHANES.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2012–2018 IRI MedProfiler Survey data.

Table A.6

Regression coefficients from percentile ranking regression adjustment

Variable	White males	White females	Hispanic males	Hispanic females	Black males	Black females	Asian males	Asian females	Other race/ ethnicity males	Other race/ ethnicity females
Age	-0.03 (0.01)	-0.01 (0.01)	-0.03 (0.02)	-0.04 (0.02)	-0.03 (0.02)	-0.06 (0.02)	-0.03 (0.02)	0.01 (0.02)	-0.07 (0.03)	0.01 (0.05)
Age^2	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
B-spline on self-reported BMI [-.15, .05)	14.31 (3.66)	11.58 (3.85)	0.3(3) (7.25)	5.34 (7.19)	10.71 (6.98)	14.69 (7.73)	8.78 (6.83)	20.60 (6.16)	(14.84) (14.63)	(1.32) (21.64)
B-spline on self-reported BMI [-0.1, 0.1)	18.83 (0.58)	17.85 (0.62)	21.36 (1.14)	20.04 (1.15)	20.08 (1.08)	19.87 (1.23)	19.20 (1.07)	15.79 (0.99)	20.54 (2.25)	19.51 (3.08)
B-spline on self-reported BMI [-.05, 0.25)	22.63 (0.28)	20.20 (0.29)	22.89 (0.54)	21.51 (0.53)	20.85 (0.51)	22.38 (0.58)	20.98 (0.52)	19.61 (0.49)	22.10 (0.96)	19.69 (1.50)
B-spline on self-reported BMI [0, 0.5)	23.36 (0.26)	21.96 (0.27)	25.30 (0.49)	23.65 (0.49)	22.90 (0.44)	25.44 (0.53)	22.40 (0.46)	19.65 (0.43)	25.88 (1.00)	21.93 (1.23)
B-spline on self-reported BMI [.05, 0.75)	26.35 (0.26)	23.89 (0.27)	27.71 (0.50)	27.09 (0.49)	25.54 (0.47)	28.87 (0.54)	24.19 (0.46)	21.30 (0.44)	26.19 (0.96)	23.78 (1.30)
B-spline on self-reported BMI [0.1, 0.9)	28.42 (0.27)	27.45 (0.27)	29.79 (0.51)	29.29 (0.50)	27.82 (0.47)	32.43 (0.53)	25.47 (0.48)	23.10 (0.45)	29.59 (0.97)	30.52 (1.34)
B-spline on self-reported BMI [0.25, 0.95)	31.89 (0.26)	31.56 (0.27)	32.51 (0.50)	34.19 (0.50)	32.38 (0.47)	37.20 (0.55)	28.34 (0.48)	26.08 (0.45)	34.38 (0.96)	31.00 (1.28)
B-spline on self-reported BMI [0.5, 1)	35.30 (0.26)	36.10 (0.27)	34.88 (0.50)	37.93 (0.49)	35.80 (0.47)	41.09 (0.53)	29.31 (0.47)	29.10 (0.44)	36.62 (0.93)	37.07 (1.38)
B-spline on self-reported BMI [0.75, 1.05)	39.43 (0.29)	43.28 (0.30)	40.23 (0.57)	43.34 (0.55)	41.50 (0.53)	49.50 (0.61)	33.28 (0.53)	32.65 (0.51)	46.25 (1.12)	45.32 (1.40)
B-spline on self-reported BMI [0.9, 1.1)	48.11 (0.58)	45.59 (0.61)	43.52 (1.12)	48.90 (1.13)	45.72 (1.08)	51.17 (1.24)	32.58 (1.04)	34.85 (0.95)	38.19 (1.99)	46.78 (3.13)
B-spline on self-reported BMI [0.95, 1.15)	95.16 (3.55)	144.64 (3.73)	103.93 (6.93)	104.04 (6.90)	149.67 (6.59)	161.71 (7.45)	124.58 (6.35)	92.61 (5.81)	199.55 (11.79)	157.41 (17.18)
Observations	3,842	3,833	2,207	2,391	2,302	2,432	1,303	1,358	400	341
R-squared	0.997	0.997	0.995	0.994	0.995	0.994	0.996	0.996	0.997	0.995

BMI = body mass index.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Regression results of measured BMI age polynomials, and cubic basis splines (b-splines) of the percentile ranks of self-reported BMI in the validation data set (NHANES) for each race/ethnicity and sex group. Sample weights are used in all calculations. Standard errors in parentheses.

Source: USDA, Economic Research Service calculations based on 2011–2018 National Health and Nutrition Examination Surveys (NHANES).

Table A.7

Mean, percentiles, and tests of distributions of reported adult BMI (kg/m^2) in MedProfiler by gender, demographic group, and adjustment method, 2017–18 only

	Mean			50th percentile			90th percentile			KS test (D-statistic)		
	Outlier methods		Predicted percentile rank method	Outlier methods		Predicted percentile rank method	Outlier methods		Predicted percentile rank method	Outlier methods		Predicted percentile rank method
	NHANES	IQR		NHANES	IQR		NHANES	IQR		NHANES	IQR	
Male												
Hispanic	29.26	28.51	29.94	28.12	27.80	29.26	37.92	35.86	37.14	0.35***	0.37***	0.03
White	26.41	26.04	25.74	25.74	25.67	25.23	32.27	31.56	30.45	0.22***	0.23***	0.02
Black	29.49	28.74	29.12	28.28	28.11	27.79	38.35	36.60	37.94	0.17***	0.18***	0.03
Asian	28.92	28.34	29.11	27.89	27.70	28.23	36.91	35.56	36.49	0.26***	0.27***	0.03
Other	28.66	27.68	29.9	27.36	27.12	28.32	37.29	35.26	40.29	0.29***	0.31***	0.06
Female												
Hispanic	28.53	28.04	30.23	27.39	27.26	29.37	37.78	36.94	38.89	0.28***	0.29***	0.04*
White	24.36	24.00	24.87	23.43	23.38	24.04	30.54	29.95	30.89	0.16***	0.17***	0.02
Black	31.09	30.23	32.62	29.53	29.17	31.63	42.90	40.34	44.24	0.18***	0.20***	0.03
Asian	28.70	28.10	29.01	27.33	26.92	27.64	38.96	37.58	39.13	0.23***	0.25***	0.03**
Other	29.17	28.45	29.78	27.40	26.95	29.42	39.54	38.08	39.49	0.31***	0.32***	0.07***

BMI = body mass index.

Note: Hispanic may be any race; race categories exclude those of Hispanic origin. Sample weights and projection factors are used in all calculations. The NHANES-outlier method excludes height and weight values that fall below or above the minimum and maximum height and weight for adults reported in NHANES. The IQR-outlier method excludes height and weight values that fall below or above the interquartile range (IQR). The predicted percentile rank method uses predictions of BMI based on a linear regression of measured BMI on percentile rankings of self-reported BMI in NHANES. The Kolmogorov-Smirnov (KS) test compares the distribution of adjusted BMIs to the distribution of measured BMIs in NHANES. Asterisk (*), double asterisk (**), and triple asterisk (***) indicate that the D-statistic is significant at the 10-, 5-, and 1-percent levels, respectively.

Source: USDA, Economic Research Service calculations based on 2017–2018 National Health and Nutrition Examination Surveys (NHANES) and the 2017 and 2018 IRI MedProfiler Surveys.