Determining Household Obesity Status Using Scanner Data

Elina T. Page
Economic Research Service
U.S. Department of Agriculture
elina.t.page@usda.gov

Sabrina K. Young
Economic Research Service
U.S. Department of Agriculture
sabrina.young@usda.gov

Megan Sweitzer

Abigail Okrent

June 15, 2021

PRELIMINARY DRAFT. PLEASE DO NOT CITE.

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting,
Austin, TX, August 1-3, 2021.

The findings and conclusions in this paper are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. The analysis, findings, and conclusions expressed in this paper also should not be attributed to IRI.

Determining Household Obesity Status Using Scanner Data

Introduction

The obesity epidemic is a global phenomenon with the rate of obesity having nearly tripled since 1975 (World Health Organization, 2018). In the United States, the current rate of obesity is higher than most other countries with over 40 percent of the adults and about 21 percent of adolescent youths classified as obese (Ogden et al., 2020). By 2030, it is projected that nearly one in two adults in the United States will be classified as obese, with nearly one in four 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 both those affected directly and for the broader society is estimated to be roughly 1.7 trillion 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 dues to lost economic productivity.

Obesity is a result of an energy imbalance, where more energy is consumed than is used by the human body. Hence, diet is the primary contributing factor to obesity, and in turn, the leading risk factor associated with death in the United States (U.S. Burden of Disease Collaborators 2018). Several studies in the nutrition literature have shown a link between body mass index (BMI), a standard measure of body fat based on height and weight, and consumer food choices. Foods that are low in essential nutrients and high in calories, such as sugar-sweetened beverages, sweets, refined cereals, and solid fats, and high in 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 needed, and the composition of foods consumed are not consistent with dietary guidelines.

Household scanner data, commercial data on household food purchases, are a rich source of information that allow researchers to link household food purchases to self-reported height and weight of household members to better study and understand the relationship between diet and obesity. Historically, household scanner data have been used to study food demand (Dong et al., 2018), variance of food choice by store format and store 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). 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 associations with specific types of foods purchased by households over time (e.g., Chen et al., 2019). However, self-reported height and weight are often misreported in survey data, and therefore it is necessary to consider the quality of these data when calculating BMI and classifying household members as normal weight, overweight, and obese.

Several studies have shown that adults tend to overreport height and underreport weight, and the degree of misreporting varies with age, body weight, and sociodemographic characteristics. Rowland (1990) found in self-reported and measured height and weight data from the National Health and Nutrition Examination Survey (NHANES) II (1976-1980) 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. Kuzmarski 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 (1988-1994). They concluded that overestimation of height by older adults may occur because greater time has elapsed since height was last measured and height decreases with age. In addition, it has also been reported that 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 have also been found to affect self-reported height and weight, as well (Stommel and Schoenborn, 2009; Wen and Kowaleski-Jones, 2012). Lastly, the degree of misreporting can also vary with how a survey is administered. In-person interviews with the expectation of having height and body weight measured in the future, as is done in NHANES, results in less misreporting of self-reported measurements compared to self-reported measurements in telephone- or web-based surveys, where there is no expectation of having height and weight measured (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 of their children. This may lead to parental reports that underestimate the measured height of children. However, in a review of studies that make direct comparisons, the mean parent-reported child height and weight were close to the corresponding measures' means, usually within one centimeter or one kilogram (Himes, 2009). Himes (2009) noted some exceptions to this; specifically, Mexican American mothers underestimated child height in the US Hispanic Health and Nutrition Examination Survey. Himes (2009) concluded that the prevalence of child obesity based on parental reports do not systematically differ from prevalence based on direct measurements.

The objective of this study is to validate and develop adjustments for self-reported height and weight measures available in household scanner data. Specifically, using the IRI Consumer Network and a subsample of households in the MedProfiler panel—a unique longitudinal dataset that tracks food purchases and body weight and height of household members over time— this study compares self-reported height and body weight in the MedProfiler to both self-reported and measured height and body weight in NHANES to better understand the quality of the IRI data. We then explore methods to correct for measurement bias by illustrating how predicting adult BMI in the IRI MedProfiler based on measured BMI and percentile rankings of self-reported BMI in NHANES works in practice. We use predicted BMI to develop body weight status categories for the household as the following: (1) normal weight households, (2) overweight households, and (3) obese households. We create these categories by defining obesity status of the household in four different ways: (a) for all household members, (b) for adults in the household, (c) for the primary shopper, and (d) for children in the household. By ensuring a reliable measure of BMI in the MedProfiler data, we hope to ultimately improve our understanding of the differences in food purchasing patterns among normal weight, overweight, and obese households.

Data Description

The primary data used in this analysis are the IRI Consumer Network household scanner data and the IRI MedProfiler survey.[1] The IRI Consumer Network household scanner data are derived from a nationwide panel of over 120,000 households that provides a detailed account of their retail food purchases, including what food products they purchased and when and where they

[1] Previous research by the U.S. Department of Agriculture has 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.

shopped. The household purchase data include product characteristic (e.g., brand, size, type) and nutrition data, which give a robust picture of the types of foods purchased by households. The Consumer Network is a nonprobability sample in which households are selected for membership into the Consumer Network through stratified quota random sampling. Under this process, households are selected based on their household characteristics to balance the panel to be representative of the U.S. population in the 48 contiguous States; in other words, they are selected to meet 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).

After households are recruited, they are asked to download a mobile application or are provided with a handheld scanner to scan the barcode or Universal Product Code (UPC) on all their purchases and transmit their purchase data on a weekly basis via the internet. A growing subsample of these 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. As a data quality check, IRI checks the consistency of weekly data reporting of panelists and identifies households in the final sample that consistently report purchases throughout the year (also called the static panel). The Consumer Network dataset includes post-stratification weights (also called projection factors) that weight the data to match the Census demographic targets, which help account for the differences between the composition of the static panel and the US general population. About half of the total participating households are included in the static panel and assigned projection factors.

The IRI MedProfiler is an opt-in survey on individuals' health and medical conditions offered to all households in the Consumer Network in October of each year (Muth et al., 2016). 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 (from 17,072 to 30,784 households), with responses received from 40,118 to 69,713 individuals. In the survey, adults over the age of 18 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 individuals), as were children under two years of age (3,022 children) in order to match to the NHANES sample. Post-stratification weights for the MedProfiler and random weight subsamples are used in all calculations in this analysis.[2]

Since reported height and body weight of household members participating in the MedProfiler are particularly important for diet quality research and there is evidence that self-reported weight suffers from downward biases and height from upward biases, we compare BMI based on self-reported height and weight in the MedProfiler to those reported and measured in the continuous NHANES for 2011-18.[3] The NHANES surveys are multi-stage probability samples of the non-institutionalized US civilian population. Both measured and self-reported height and weight information are collected in NHANES for the same individuals. Measured height and weight are collected in physical examination centers by trained health technicians, and self-reported height and weight are collected during interviews. The final 2011-18 NHANES sample consisted of 22,716 adults aged 16 years and older and 10,693 children aged 2 to 15

[2] Since the projection factors scale the IRI MedProfiler-random weight sample to be consistent with demographic targets in the US population for a year, we divide the projection factors by a factor of 7 to produce 7-year estimates over the six cycles of IRI MedProfiler.

[3] BMI is calculated as body mass (weight) divided by the square of body height.

years of age. Participants missing measured weight or height variables (2,168 individuals) were excluded. Only adults aged 16 years and older were asked to report height and weight in interviews, and of those, 764 were excluded because they refused, didn't know or were missing height or weight information. The sample weights for the examination component were used in estimation.[4]

Validation of BMI in Household Scanner Data

We compare the BMI distributions between the MedProfiler and NHANES separately for children aged 2-15 years and adults aged 16 years and older. Since children and youths are constantly growing, and overweight and obese cutoffs vary by age and gender, we compare BMI distributions for children between the two datasets by each gender and year of age separately. For adults, we examine the BMI distributions based on each dataset by gender and demographic profile.[5] Also, for adults aged 16 and older, NHANES provides information on both measured and self-reported height and weight, which we use in adjusting self-reported BMI in the IRI MedProfiler. This information was unavailable for children.

We compare the demographics, average weight, height, BMI, and proportion of each dataset that were underweight, normal weight, overweight and obese across datasets (table 1). Adults aged 16 years and above in the MedProfiler are more non-Hispanic white (69 percent versus 65 percent) and less male (45 percent versus 49 percent) compared to their adult counterparts in NHANES. The composition of the child samples between the two sources (ages 2-15 years) differs much more. The NHANES child sample contains 52 percent non-Hispanic

[4]The two-year sample weights were combined to produce 6-year estimates over three cycles of NHANES by dividing the two-year weights by a factor of 3 (Chen et al., 2018).
[5] We intended to extend the analysis of children to demographic groups, but the count of children in each gender-age cell in NHANES was too small to perform this analysis.

whites whereas the entire MedProfiler child sample contains 64 percent. About 24 and 14 percent of the NHANES child sample are Hispanic and non-Hispanic black, respectively, compared to 19 and 11 percent of the MedProfiler child sample. The differences in gender composition between the two datasets is relatively small (1 percentage point). The adult MedProfiler sample likely aligns more closely with the adult NHANES because the projection factors are constructed based on matching the household heads demographic characteristics to Census demographic targets, but no such weighting is used for children in the MedProfiler.

Table 1. Composition of NHANES and IRI MedProfiler Samples

	Ages 2-15		**Ages 16+**	
	NHANES	**MedProfiler**	**NHANES**	**MedProfiler**
Sample size	10,693	45,646	22,716	338,681
Hispanic (%)	24	19	14	13
Non-Hispanic White (%)	52	64	65	69
Non-Hispanic Black (%)	14	11	12	11
Non-Hispanic Asian (%)	05	05	06	04
Other (%)	06	02	04	02
Male (%)	51	52	49	45
Actual measurements				
Mean BMI (kg/m^2)	19.28	--	28.97	--
Proportion underweight (%)	03	--	02	--
Proportion normal weight (%)	63	--	29	--
Proportion overweight (%)	16	--	32	--
Proportion obese (%)	18	--	37	--
Self- or parent-reported measurements				
Mean BMI (kg/m^2)	--	21.02	28.24	28.60
Proportion underweight (%)	--	13	02	02
Proportion normal weight (%)	--	52	33	31
Proportion overweight (%)	--	14	33	32
Proportion obese (%)	--	21	33	34

Source: USDA Economic Research Service calculations based on 2011-18 National Health and Nutrition Examination Surveys (NHANES) and the 2012-18 IRI MedProfiler Survey.
Notes: Sample weights and projection factors used in all calculations. For adults, a BMI below 18.5 is underweight, between 18.5 and 24.9 is normal weight, between 25.0 and 29.9 is overweight, and 30.0 and above is obese. For children, obesity classifications also consider age and gender (see https://www.cdc.gov/obesity/childhood/defining.html).

Like many previous studies, self-reported height is overreported and weight is underreported in NHANES compared to measured height and weight values in NHANES. For the 2011-18 NHANES, the average self-reported height for adults 16 years and older is 1.70 meters, which is 0.01 meters more than that measured height. Self-reported weight is 0.78 kilograms less than its measured counterpart in NHANES. Hence, BMI based on measured height and weight in NHANES was 0.73 kg/m^2 more than BMI based on self-reported height and weight, and the population characterized as being obese was 4 percentage points less. We also found that not only was average self-reported height in the MedProfiler adult sample higher compared to measured height in NHANES, but also average self-reported weight was larger in the MedProfiler as well. Similar to average BMI based on self-reported height and weight in NHANES, BMI based on MedProfiler is less than average BMI based on measured height and weight in NHANES (28.60 versus 28.97 kg/m^2). Hence, less of the MedProfiler adult sample is categorized as obese compared to categorizations based on measured NHANES BMI.

Because NHANES only contains measured height and weight for children, we compare the measured NHANES height and weight with parent-reported height and weight in the MedProfiler. Average measured body height (1.34 m) and weight (37.66 kg) is quite a bit larger compared to parent-reported height (1.30 m) and weight (33.89 kg). BMI based on measured height and weight for children and youths is lower (19.28 kg/m^2) on average compared to parent-reported BMI (21.02 kg/m^2).

Figure 1 compares the densities of adult BMI by dataset, race, and gender, and shows some differences between the distributions. The dispersion of the data (i.e., standard deviation) and skewness (i.e., symmetry of the distribution) between the measured and self-reported BMI in NHANES and IRI MedProfiler are similar. However, across all gender and demographic profiles, the median and 90th percentiles for self-reported BMI in the IRI MedProfiler are generally less than those using measured BMI in NHANES (table 2). Additionally, the BMI distribution based on the IRI MedProfiler has more extreme BMI values (i.e., larger kurtosis) compared to both self-reported and measured BMI in NHANES.

Figure 2 compares the relationship between age and BMI between the two datasets (measured NHANES and the IRI MedProfiler). Both datasets show the expected pattern between age and BMI of an inverse U-shaped relationship across all demographic groups, with non-Hispanic Asian males and females having the lowest BMI across all ages. However, NHANES shows that non-Hispanic white males and females have BMIs lower than non-Hispanic black and Hispanic males and females for most ages, but the IRI MedProfiler does not show similar results.

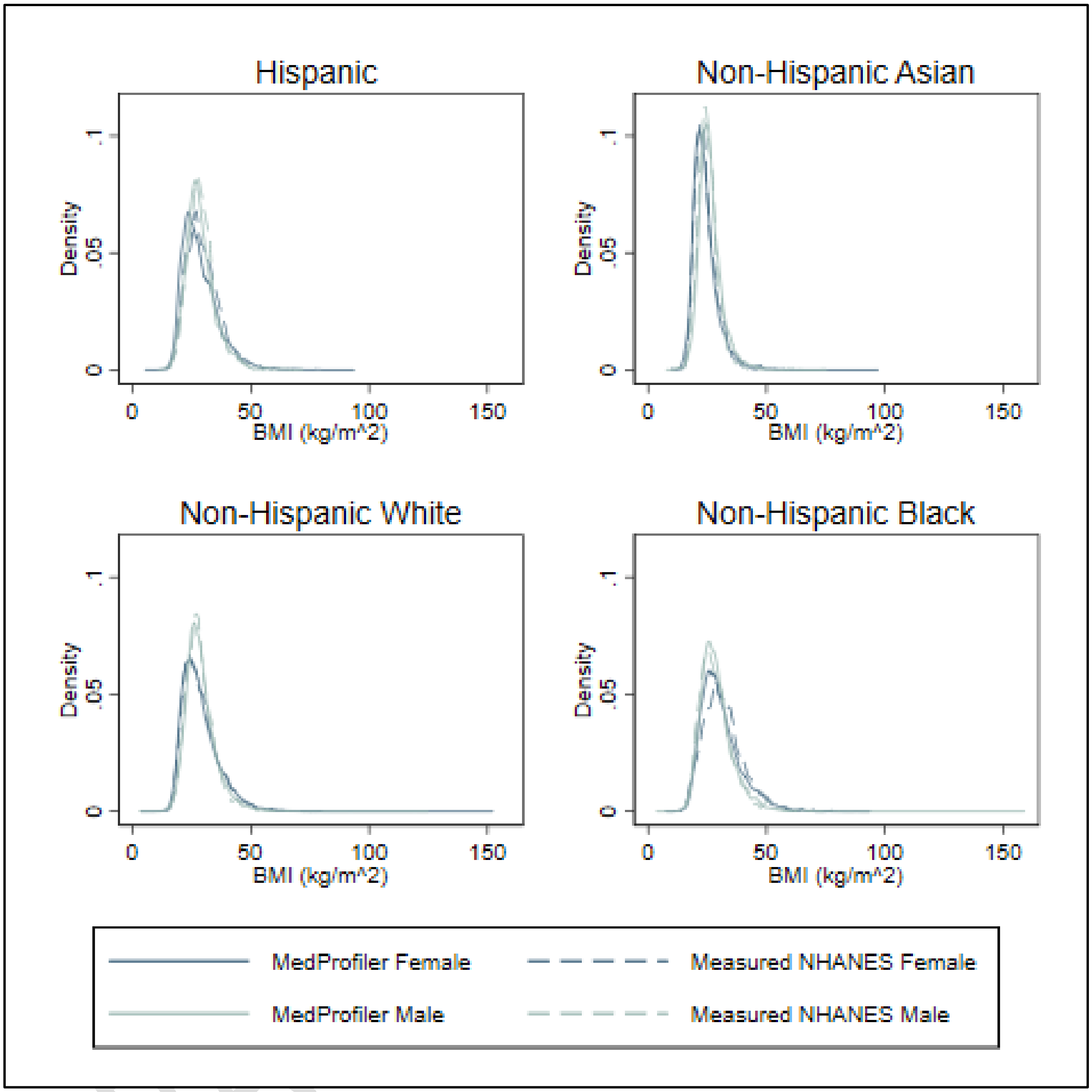

Figure 1. Self-Reported and Measured BMI for Adults (16+ Years Old)
Source: USDA Economic Research Service calculations based on epanechnikov kernel density estimations (half-width of kernel equal to 3) and the 2011-18 National Health and Nutrition Examination Surveys (NHANES) and the 2012-18 IRI MedProfiler Survey data.
Notes: Sample weights and projection factors used in all calculations.

Table 2. Comparison of Distributions of Measured and Self-Reported Adult BMI by Race and Gender Across Datasets (kg/m^2)

	Median			90th Percentile		
	NHANES (Measured)	**NHANES (Self-Reported)**	**MedProfiler**	**NHANES (Measured)**	**NHANES (Self-Reported)**	**MedProfiler**
Male						
Hispanic	28.80	28.39	27.49	36.80	36.21	36.91
Non-Hispanic White	27.90	27.29	27.60	36.40	35.19	36.80
Non-Hispanic Black	27.60	27.34	27.76	38.10	36.95	37.75
Non-Hispanic Asian	25.00	24.82	25.10	30.80	29.89	32.25
Other	28.40	27.77	27.31	39.50	38.34	36.51
Female						
Hispanic	28.90	27.83	26.82	39.50	38.41	38.60
Non-Hispanic White	27.40	26.49	26.95	39.30	37.80	38.96
Non-Hispanic Black	30.70	29.56	28.97	43.30	41.09	41.88
Non-Hispanic Asian	23.70	23.05	23.04	31.30	30.14	30.66
Other	28.80	27.85	27.45	40.60	38.64	41.00

Source: USDA Economic Research Service calculations based on 2011-18 National Health and Nutrition Examination Surveys (NHANES) and the 2012-18 IRI MedProfiler Survey data.
Notes: Sample weights and projection factors used in all calculations.

Table 2 (Continued). Comparison of Distributions of Measured and Self-Reported Adult BMI by Race and Gender Across Datasets (kg/m^2)

	Standard Deviation			Skewness			Kurtosis		
	NHANES (Measured)	NHANES (Self-Reported)	MedProfiler	NHANES (Measured)	NHANES (Self-Reported)	MedProfiler	NHANES (Measured)	NHANES (Self-Reported)	MedProfiler
Male									
Hispanic	5.90	5.58	6.44	0.89	1.28	1.28	4.83	4.49	6.52
Non-Hispanic White	6.16	5.73	6.40	1.06	1.37	1.37	5.07	5.18	8.11
Non-Hispanic Black	7.19	6.41	7.04	1.32	1.70	1.70	6.59	5.22	12.96
Non-Hispanic Asian	4.31	4.10	4.92	1.39	1.30	1.30	9.54	11.98	7.55
Other	7.49	6.96	6.56	1.22	1.44	1.44	6.01	5.83	8.57
Female									
Hispanic	7.22	6.93	7.66	0.88	1.33	1.33	4.02	4.22	6.04
Non-Hispanic White	7.59	7.15	7.72	1.20	1.33	1.33	5.19	5.55	6.15
Non-Hispanic Black	8.72	7.95	8.27	0.99	1.16	1.16	4.70	4.81	5.04
Non-Hispanic Asian	4.82	4.49	5.35	1.10	1.86	1.86	4.82	4.93	11.06
Other	8.30	8.00	8.53	1.46	1.34	1.34	7.23	6.74	5.76

Source: USDA Economic Research Service calculations based on 2011-18 National Health and Nutrition Examination Surveys (NHANES) and the 2012-18 IRI MedProfiler Survey.
Notes: Sample weights and projection factors used in all calculations.

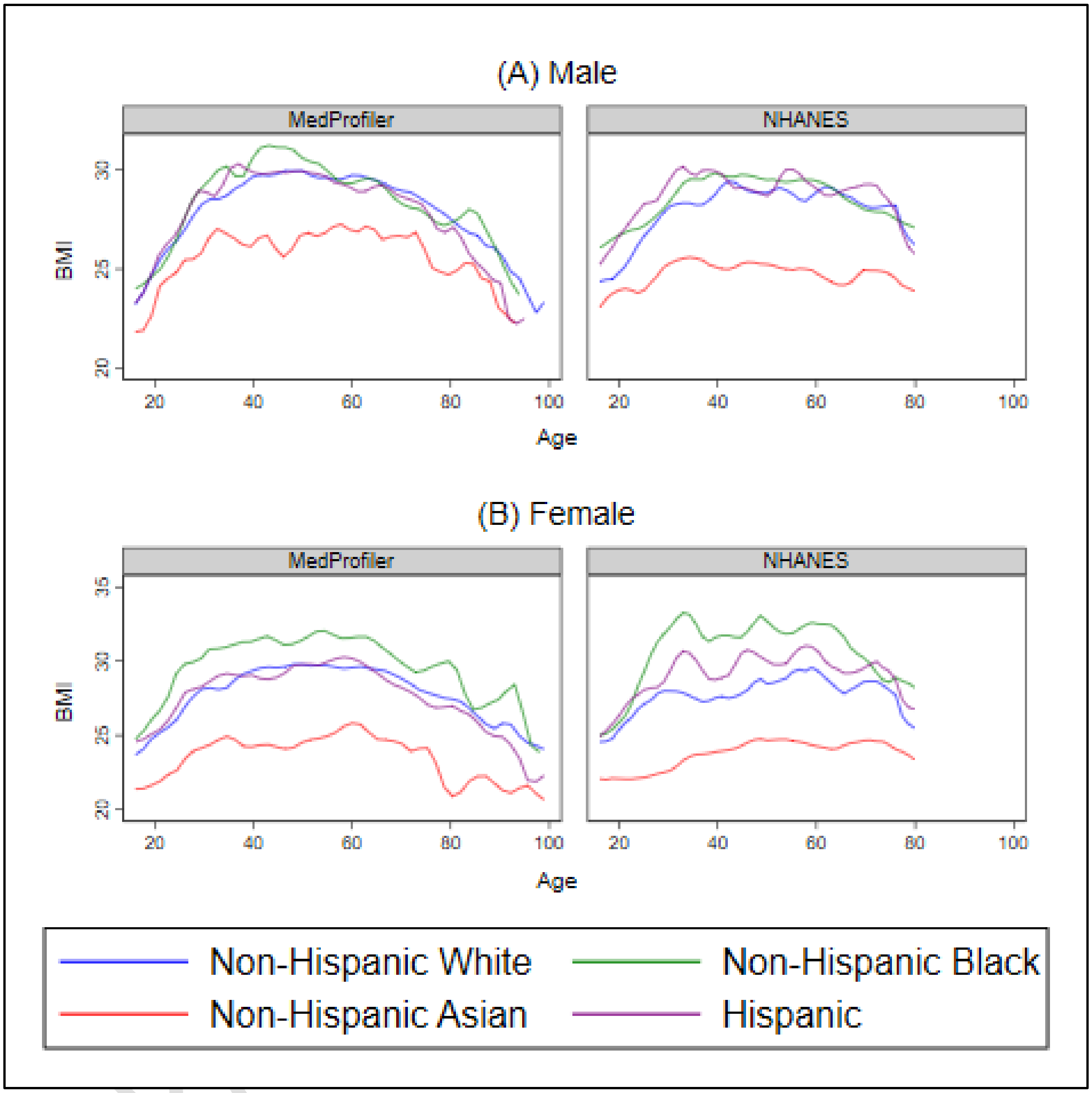

Figure 2. Relationship between Age and BMI for Adults (Age 16+) by Gender and Demographic Group

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

Notes: Sample weights and projection factors used in all calculations.

Differences in the BMI distributions between the two datasets vary by age and gender for the child samples as well. For male and female children, parent-reported median BMI in the IRI MedProfiler are consistently lower than NHANES starting at age 6 and the 90th percentile is consistently lower starting at age 9 (table 3). The distribution of parent-reported BMIs in the IRI MedProfiler sample of children is much more skewed to the right with fatter tails compared to that using NHANES across most of the age-gender groups. Also, the dispersion of the IRI MedProfiler BMI distribution tends to decline with age.

Like the adult samples, the relationship between age and BMI for children in the self-reported IRI MedProfiler and measured NHANES (figure 3) are similar. After age 5, BMI increases with age with non-Hispanic Asian children having the lowest BMI, followed by non-Hispanic white children.

Table 3. Comparison of Distributions of Measured and Self-Reported Child BMI by Age and Gender Across Datasets (kg/m^2)

	Median		90th Percentile		Standard Deviation	
	NHANES (Measured)	MedProfiler	NHANES (Measured)	MedProfiler	NHANES (Measured)	MedProfiler
Male						
2	16.7	17.1	18.6	27.1	1.5	1223.3
3	16.2	16.3	18.3	21.9	2.1	6.1
4	15.9	16.2	18.4	21.2	1.8	5.1
5	16.1	15.9	18.6	21.7	2.1	5.7
6	16.0	15.8	19.7	22.6	2.7	5.2
7	16.4	16.1	22.2	22.8	3.2	4.5
8	16.9	16.8	22.9	24.2	3.6	4.9
9	17.4	17.4	24.4	23.8	3.9	4.7
10	19.0	18.2	26.1	25.0	4.6	4.8
11	19.7	18.8	27.1	25.3	4.9	12.4
12	19.3	19.5	27.7	26.9	4.7	4.9
13	20.6	20.1	30.9	27.4	5.6	4.8
14	21.6	20.5	30.2	28.3	5.4	13.9
15	22.3	21.4	32.5	29.0	6.1	5.0
Female						
2	16.4	16.8	18.2	26.9	1.4	7.5
3	15.9	16.3	17.9	22.6	1.6	6.4
4	16.1	15.9	18.7	21.7	2.0	4.7
5	15.8	15.9	18.9	21.9	2.1	5.2
6	16.0	15.9	21.0	21.7	3.1	5.1
7	16.4	16.2	21.9	22.6	3.3	4.6
8	17.2	16.8	23.2	24.0	3.8	5.3
9	17.8	17.6	25.0	24.4	4.2	5.1
10	19.1	18.3	25.3	24.7	4.3	4.8
11	19.5	19.1	28.4	26.9	5.1	5.9
12	21.3	19.8	29.5	26.6	5.5	4.7
13	20.7	20.6	30.3	28.2	5.4	5.0
14	22.4	21.2	30.1	29.8	5.2	5.6
15	22.6	21.7	32.6	30.9	5.8	5.7

Source: USDA Economic Research Service calculations based on 2011-18 National Health and Nutrition Examination Surveys (NHANES) and the 2012-18 IRI MedProfiler Survey.
Notes: Sample weights and projection factors used in all calculations.

Table 3 (Continued). Comparison of Distributions of Measured and Self-Reported Child BMI by Age and Gender Across Datasets (kg/m²)

	Skewness		Kurtosis	
	NHANES (Measured)	MedProfiler	NHANES (Measured)	MedProfiler
Male				
2	1.2	21.7	7.2	473.5
3	3.1	5.6	18.8	59.6
4	1.8	3.6	8.2	24.0
5	1.6	7.6	6.5	147.7
6	2.8	3.0	16.3	19.6
7	1.8	2.2	7.9	13.3
8	1.5	1.8	6.1	10.3
9	1.5	1.7	5.7	10.6
10	1.4	1.6	5.9	7.7
11	1.0	22.2	4.1	619.0
12	1.2	1.2	4.3	5.7
13	1.2	1.2	4.3	5.3
14	1.3	41.7	5.1	1972.1
15	1.2	1.3	4.6	5.7
Female				
2	0.8	3.9	4.8	30.7
3	3.5	7.2	32.3	102.6
4	1.3	3.3	5.4	26.4
5	1.6	4.1	6.1	38.0
6	1.6	2.7	6.3	14.9
7	1.4	1.8	5.4	8.9
8	1.8	2.5	8.9	14.5
9	1.3	2.8	4.8	24.4
10	1.1	1.7	5.2	8.9
11	1.1	3.6	4.0	30.8
12	1.0	1.3	4.1	6.5
13	1.4	1.6	5.3	7.4
14	1.2	2.4	4.6	19.3
15	1.4	1.8	5.3	9.1

Source: USDA Economic Research Service calculations based on 2011-18 National Health and Nutrition Examination Surveys (NHANES) and the 2012-18 IRI MedProfiler Survey.
Notes: Sample weights and projection factors used in all calculations.

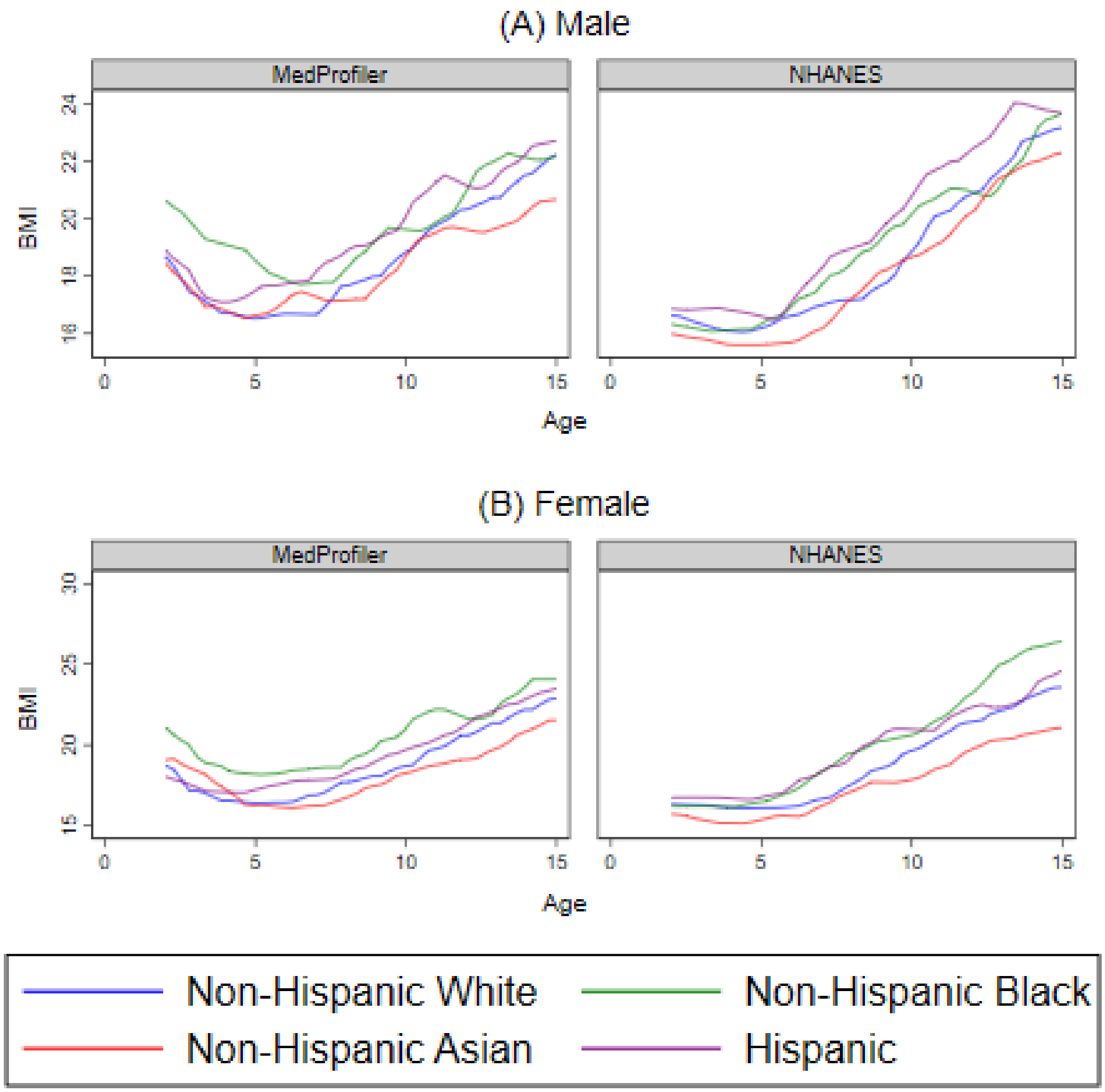

Figure 3. Relationship between Age and BMI for Children (Age 2-15) by Gender and Demographic Group

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

Notes: Sample weights and projection factors used in all calculations. Two implausible outliers (BMI>500) were excluded for ease of viewing overall trends.

Overall, misreporting of height and weight, and hence BMI, seems to be less prevalent in the adult sample compared to the child sample in the IRI MedProfiler. The BMI distributions for male non-Hispanic white and black adults and female non-Hispanic white and Asian adults in the IRI MedProfiler align well with the measured BMI in NHANES in terms of median and 90th percentiles, skewness and standard deviation. However, the BMI distributions for the other demographic adults are quite different between the reported and measured IRI MedProfiler and NHANES, respectively, and kurtosis is larger for all demographic-gender groups as well. The differences in medians of each age-gender distribution between the measured NHANES and the parent-reported IRI MedProfiler are generally less than 1kg/m^2 with the exception of male children aged 14 and female children aged 12 and 14. However, the differences in the 90th percentiles are quite large. Parents tend to overreport BMI for children 8 years of age or less, and overreport BMI for children older than 8 years of age.

Methods for Reducing Measurement Bias

Future work will explore methods to remove outliers and adjust BMI to reduce measurement bias in the self-reported IRI MedProfiler data. First, we will develop criteria for eliminating BMI outliers for adult and children distributions in the IRI MedProfiler data. Second, we will correct misreporting of BMI based on self- and parent-reported height and weight using self-reported and measured BMI in NHANES for adults aged 16 and older. We will remove outliers two ways to examine how they affect the BMI distributions: (1) based on the interquartile range (IQR-outlier method) (e.g., Rousseeuw and Croux, 1993), and (2) based on the minimum and maximum measured BMI values reported in NHANES (NHANES-outlier method) (e.g., Freedman et al., 2015). To correct for correct for measurement bias in BMI based on self-reported height and weight in adults, we plan to follow Courtemanche et al. (2015).

Courtemanche et al. developed correction method for self-reported weight and height that relies the percentile rank of reported values in their respective distributions rather than reported values. Hence, their 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 resemble the rankings based on actual measures in both datasets, and both datasets represent the same population (e.g., nationally representative samples).

Conclusion and Future Research

The IRI 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 are a promising tool to study the links between food purchases and health outcomes, but little information was available about the quality of these self-reported data. This analysis compared BMI from self-reported height and weight in the MedProfiler data to their measured counterparts in NHANES to assess the value of these data for food and health policy research. We find that male non-Hispanic white and black adults and female non-Hispanic white and Asian adults in the IRI MedProfiler align well with the measured BMI in NHANES, while other groups and children show some significant differences.

Future work will explore several methods to adjust BMI distributions to reduce the measurement bias introduced from self-reported data. We will further explore how predicting adult BMI in the IRI MedProfiler based on measured BMI and percentile rankings of self-reported BMI in NHANES works in practice. We will use predicted BMI to develop body weight status categories for the household as the following: (1) normal weight households, (2) overweight households, and (3) obese households. We will create these categories by defining obesity status of the household in four different ways: (a) for all household members, (b) for adults in the household, (c) for the primary shopper, and (d) for children in the household. Using fruit and vegetable expenditures as an example, we plan to compare expenditures by body weight classification in the above four methods.

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* 54:143-148.

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.

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):412–419.

Courtemanche, C., J.C. Pinkston, and J. Stewart. 2015. "Adjusting Body Mass for Measurement Error with Invalid Validation Data." *Economics and Human Biology*, 19:275-293.

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):122-136.

Freedman, D.S., H.G. Lawman, A.C. Skinner, L.C. McGuire, D.B. Allison, and C.L. Ogden. 2015. "Validity of the WHO cutoffs for biologically implausible value of weight, height, and BMI in children and adolescents in NHANES from 1999 through 2012." *Ameican Journal of Clinical Ntrition*, 102(5):1000-1006.

Himes, J.H. 2009. "Challenges of Accurately Measuring and Using BMI and Other Indicators of Obesity in Children," *Pediatrics*, 124(Supplement 1):s3-s22.

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." *American Dietetic Association*, 101(1):28-34.

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, C.M. Hales. 2020. "Trends in Obesity Prevalence by Race and Hispanic Origin – 1999-200 to 2017-2018." *Journal of the American Medical Association (JAMA)*, 324(12):1208-1210.

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

Rousseeuw, P.J., and C. Croux. 1991. "Alternatives to the median absolute deviation." *Journal of the American Statistical Association* 88(424):1273-283.

Rowland, M.L. 1990. "Self-Reported Weight and Height." *The American Journal of Clinical Nutrition*, 52(6):1125-1133.

Spencer, E.A., P.N. Appleby, G.K. Davey, and T.J. Key. 2001. "Validity of Self-Reported Height and Weight in 4808 EPIC-Oxford Participants." *Public Health Nutrition*, 5(4): 561-565.

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 US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among U.S. States." *Journal of the American Medical Association (JAMA)*, 319(14):1444-1472.

Volpe, R., and A. Okrent. 2012. *Assessing the Healthfulness of Consumers' Grocery Purchases*, Economic Information Bulletin 102, U.S. Department of Agriculture, Economic Research Service.

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

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):2440-2450.

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.*

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.