2-2-2013

Adjusting Body Mass for Measurement Error with Invalid Validation Data

Charles J. Courtemanche
Georgia State University, ccourtemanche@gsu.edu

Joshua Pinkston
University of Louisville, josh.pinkston@louisville.edu

Jay Stewart
Bureau of Labor Statistics & IZA

Recommended Citation

Courtemanche, Charles J.; Pinkston, Joshua; and Stewart, Jay, "Adjusting Body Mass for Measurement Error with Invalid Validation Data" (2013). *UWRG Working Papers*. 70.
https://scholarworks.gsu.edu/uwrg_workingpapers/70

Adjusting Body Mass for Measurement Error with Invalid Validation Data

Charles Courtemanche
Georgia State University & NBER

Joshua C. Pinkston
University of Louisville

Jay Stewart
Bureau of Labor Statistics & IZA

GeorgiaState University® | ANDREW YOUNG SCHOOL

OF POLICY STUDIES

Adjusting Body Mass for Measurement Error

with Invalid Validation Data[1]

Charles Courtemanche
Georgia State University & NBER

Joshua C. Pinkston
University of Louisville

Jay Stewart
Bureau of Labor Statistics & IZA

February 14, 2013

Abstract

We propose a new method for using validation data to correct self-reported weight and height in surveys that do not weigh and measure respondents. The standard correction from prior research regresses actual measures on reported values using an external validation dataset, and then uses the estimated coefficients to predict actual measures in the primary dataset. This approach requires the strong assumption that the expectations of actual weight and height conditional on the reported values are the same in both datasets. In contrast, we use percentile ranks rather than levels of reported weight and height. Our approach requires the much weaker assumption that the conditional expectations of actual measures are increasing in reported values in both samples, making our correction more robust to differences in measurement error across surveys. We then examine three nationally representative datasets and confirm that misreporting is sensitive to differences in survey context such as data collection mode. When we compare predicted BMI distributions using the two approaches, we find that the standard correction is biased by differences in misreporting while our correction is not. Finally, we present several examples that demonstrate the potential importance of our correction for future econometric analyses and estimates of obesity rates.

---

1 Introduction

Obesity, defined as having a body mass index (BMI) of at least 30, leads to heart disease, diabetes, high blood pressure, stroke, and other health problems (Strum, 2002).[1] The obesity rate among U.S. adults rose from 13% in 1960 to 34% in 2008, prompting a large literature examining the causes and consequences of this trend (Flegal et al., 1998; National Center of Health Statistics, 2008).

Appropriate measurement of weight and height has long been a problematic issue for researchers in this literature. Medically measured weight and height are obviously ideal, but they are expensive to collect in large samples due to the need for trained personnel to have in-person contact with respondents. For this reason, only one large-scale U.S. government health dataset – the National Health and Nutrition Examination Surveys (NHANES) – contains measured weights and heights. The NHANES has several limitations that render it viable for only a small subset of the topics of interest to obesity researchers. Though large enough to produce national-level descriptive statistics, it is often too small for more sophisticated estimation. Moreover, the NHANES consists of repeated cross-sections so it does not allow for the use of panel data methods. Finally, although the NHANES contains excellent health information, it includes only a limited number of economic and demographic variables.

Other datasets include respondents' self-reports of weight and height, often obtained through telephone surveys (e.g. Behavioral Risk Factor Surveillance System (BRFSS)). Such an approach permits a much larger sample size and broader geographic coverage, but is limited by the fact that self-reports are often subject to considerable measurement error. Some respondents

---

[1] BMI=weight in kilograms divided by height in squared meters.

may not know their current weight, while others might be dishonest about their weight and height in an effort to adhere to social norms. Cawley (2002) finds that underweight people tend to overreport their weight while those who are heavier tend to underreport. Overreporting among the underweight is greater for men, while underreporting among the overweight is greater for women. For both sexes, the distribution of self-reported weight is therefore more compressed than the actual weight distribution. Rowland (1990) finds similar results for weight and also documents a tendency to exaggerate height that is more pronounced among the overweight. These results imply that obesity rates computed from self-reported weight and height will be understated. Moreover, the systematic, non-classical nature of the measurement error suggests that bias in regression estimates is possible regardless of whether BMI is an independent or dependent variable, and that the direction of the bias is unclear.

In two concurrent papers, Cawley (2002; 2004) corrects for the misreporting of height and weight using a procedure that was developed in the statistics literature. In a dilemma common in obesity research, the NHANES is not a suitable dataset for the topic of either paper.[2] In both studies, Cawley uses the 1979 cohort of the National Longitudinal Survey of Youth, which contains self-reported weight and height, and attempts to correct these measures using the relationship between self-reports and measured values in the NHANES. For each race and gender group, Cawley regresses actual weight and height on the corresponding self-reports and their squares in NHANES, and then uses the resulting regression estimates to predict the NLSY79 respondents' actual weights and heights.

---

[2] The first paper, Cawley (2002), tests for rational addiction in caloric intake, which requires panel data. The second, Cawley (2004), examines the impact of obesity on wages, and wages are not available in the NHANES.

Cawley's correction for self-reported weight and height has since become commonly used in the economics of obesity literature. Several recent papers have used this correction when studying the impacts of obesity on labor market outcomes (e.g. Cawley and Danzinger, 2005; Gregory, 2010; Gregory and Ruhm, 2011; Majumder, 2013). This correction has also been used in a number of papers that examine potential economic determinants of obesity.[3]

Unfortunately, the standard validation method of Cawley (2002, 2004) is not appropriate if the amount or type of measurement error differs between the primary and validation samples. As Han et al. (2009) point out, NHANES respondents expect to be weighed and measured when they report their height and weight, while respondents in the NLSY and other commonly used datasets do not. Furthermore, Pinkston (2013) notes that even being interviewed by phone instead of in person affects the self-reported values of respondents in the NLSY cohorts.

Our paper develops an alternative correction for self-reported weight and height that relies on weaker assumptions about the relationship between true and reported values in the primary and validation datasets. Instead of relying on the reported values themselves, we predict actual measures using the percentile rank of reported values in their respective distributions. Our 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 resembles the rankings based on actual measures in both datasets.

---

[3] Potential determinates of obesity studied using Cawley's correction include age (Baum and Ruhm, 2009), income (Cawley et al., 2010), unemployment rate (Ruhm, 2005), childhood socioeconomic status (Baum and Ruhm, 2009), food prices (Lakdawalla and Philipson, 2002; Chou et al., 2004; Courtemanche et al., forthcoming; Goldman et al., 2011), cigarette prices (Chou et al., 2004; Baum, 2009), alcohol prices (Chou et al., 2004), food stamps (Fan, 2010; Baum, 2011), restaurant density (Chou et al., 2004), on-the-job physical activity (Lakdawalla and Philipson, 2002), smoking bans (Chou et al., 2004), urban sprawl (Plantinga and Bernell, 2007; Eid et al., 2008), and time preference (Courtemanche et al., forthcoming).

Using data from the Behavioral Risk Factor Surveillance System and the American Time Use Survey, we show that our percentile rank method produces predicted values of BMI that are consistent across datasets, and consistent with the distribution of actual BMI in the population. We also confirm that the standard approach in the literature is not appropriate in these datasets because it is sensitive to differences in misreporting across samples.

Finally, we illustrate how the corrections can influence regressions coefficients and estimates of the prevalence of obesity. Using both the BRFSS and ATUS, we consider basic regressions that include BMI or obesity as either a dependent or an independent variable. In each case, we compare estimates with our adjustment to analogous estimates that use either no correction for measurement error or the standard correction. While using our correction generally does not affect the signs of coefficient estimates or statistical significance, it can lead to important differences in the magnitudes of the estimates. We then revisit the Centers for Disease Control's (CDC's) well-known map of obesity rates by state and demonstrate that correcting for measurement error dramatically increases the estimated prevalence for most states.

2 The Problem of Transportability and an Alternative Approach

For the sake of simplicity, our discussions in this section will consider the regression of some dependent variable on an independent variable that is measured with error. Both the standard method and our proposed method extend to cases in which the dependent variable is measured with error under similar assumptions.[4] In what follows, we will note any changes in assumptions required by such extensions.

---

[4] L&S develop their approach to correct measurement error in either the dependent or independent variables of a nonlinear regression. See Chen et al. (2005) among others for applications that use fewer parametric assumptions.

Let b denote the true measures of height or weight in the population, and $\tilde{b}_j$ denote the

reported value in sample j, where $j = P, V$ indicates the primary or validation dataset. The

reported values, $\tilde{b}_j$, are allowed to have arbitrary (potentially nonclassical) measurement error.

The standard validation approach used by Cawley (2002, 2004) is based on work by Lee

and Sepanski (1995) (L&S in what follows) and others in the statistics literature.[5] We can distill

two conditions from this literature that must be met when validation data are used to correct for

measurement error:

C1. There must be a *surrogate* for b. A variable, b_j^s, is a surrogate for b if the distribution

of y given (b, b_j^s) is the same as the distribution of y given b.[6] In cases where b is a

dependent variable that is measured with error, b_j^s is a surrogate if its distribution

depends only on the true response (see Carroll et al., 2006).

C2. The surrogate, b_j^s, must satisfy some form of *transportability* across datasets.

Transportability is usually described as the distribution of b conditional on b_j^s being

the same in both datasets; however, L&S use a weaker form of transportability,

which requires $E(b|b_P^s) = E(b|b_V^s)$.[7]

The first condition simply states that a surrogate for b contains no information about the

dependent variable that is not also contained in b (and possibly other observed covariates). This

condition is easily satisfied by the reported values, $\tilde{b}_j$.

---

[5] See Bound et al. (2001) for a brief survey of this work, and Carroll et al. (2006) for more depth.
[6] In addition to L&S, see Carroll et al. (2006).
[7] See Bound et al. (2001) or Carroll et al. (2006) for examples of the stronger version of transportability. Strictly speaking, L&S assume that the expectation of y conditional on b_j^s is the same in both datasets, but that reduces to $E(b|b_P^s) = E(b|b_V^s)$ in the current context.

The second condition, transportability, is essential if the procedure used to generate predicted values in the validation dataset is to be applied to the primary dataset. C2 requires that researchers make additional assumptions about the characteristics of the primary and validation datasets. Carroll et al. (2006) warn that transportability may not be satisfied when the validation data are drawn from external sources. They emphasize that validation data are ideally drawn from a random subsample of the primary data. Bound et al. (2001) point out that measurement error in survey data often varies with the context of the survey, which is potentially an issue when the validation data are drawn from an external source.

Bound et al. (2001) also note that transportability requires that the primary and validation data are representative of the same population. As in previous work on obesity, we use datasets that are weighted to be nationally representative, which implies that the distribution of b, $F(b)$, does not vary across samples. In the next section, we will use this implication to demonstrate that transportability is not satisfied when using the standard approach.

The rest of this section compares the standard validation approach used by Cawley (2002, 2004) and our proposed alternative. In section 2.1, we focus on the assumptions required for the standard approach to satisfy transportability, and argue that they are not likely to hold. In section 2.2, we propose an alternative surrogate that satisfies the transportability condition under weaker assumptions. In contrast to the standard approach, the method we develop is robust to differences in misreporting across samples.

2.1 The Standard Validation Method

Consider the regression of some dependent variable y on b and other covariates, but assume we only observe $\tilde{b}_P$ in the primary data set. Following the previous literature on obesity, we would estimate

$$b = \eta(\tilde{b}_V) + \epsilon \qquad (1)$$

using the validation sample, and then use

$$\hat{b} = \hat{\eta}(\tilde{b}_P)$$

in place of b as an independent variable in our primary sample.[8]

This approach assumes that $\tilde{b}_j$ is a surrogate for b, and that it is transportable. In this case, transportability is satisfied under the assumption that $E(b|\tilde{b}_P) = E(b|\tilde{b}_V)$.[9] This assumption is violated if measurement error in $\tilde{b}_j$ varies across samples, which is likely when NHANES data are treated as a validation sample for data from telephone surveys. As we noted above, respondents are more likely to misreport their height and weight in telephone surveys than in in-person surveys and misreport to a greater extent.

2.2 An Alternative Method Based on Percentile Rank

Fortunately, the theory supporting the use of validation data does not require us to directly use the reported measures, $\tilde{b}_j$, when correcting for measurement error in b. Any surrogate for b that satisfies transportability can be used.

---

[8] In Cawley (2004) and other papers in the obesity literature, the dependent variable is actually regressed on a nonlinear function of the predicted values, BMI. As we discuss later, L&S argue that it would be preferable to predict the nonlinear function directly.

[9] This is consistent with the weaker form of transportability used by L&S.

As before, we assume that the $\tilde{b}_j$ are surrogates for b. Following the previous literature, $\tilde{b}_j$ are functions of b and a random error term that is not correlated with y.[10] This implies that the percentile rank of $\tilde{b}_j$, given by the distribution function $G_j(\tilde{b}_j)$, is a function of b and the same random error. Therefore, $G_j(\tilde{b}_j)$ is also a surrogate for the true value, b, satisfying C1.

Our approach uses the percentile rank of the report, $\tilde{b}_j$, to generate predicted values of b. The advantage of this approach is that it requires a much weaker assumption about the relationship between $\tilde{b}_j$ and the true values, b. Specifically, we assume:

A1. The expected value of the true measure conditional on the reported value is monotonically increasing in the reported value; i.e., $\tilde{b}'' > \tilde{b}'$, implies that $E(b|\tilde{b}'') > E(b|\tilde{b}')$.[11]

In other words, A1 says that, given any two people who report their weight, the person who reports the higher weight is expected to actually weigh more. Assumption A1 is unaffected by differences in measurement error across samples, as long as reported values still allow a relevant (expected) ranking of respondents in each sample. Furthermore, if A1 does not hold, then it is not clear to us that respondent reports convey any useful information about actual height and weight. Finally, note that the monotonicity described by A1 is testable in the validation sample but not in the primary samples.[12]

---

[10] Following equation (1), $\tilde{b}_v = \eta^{-1}(b - \epsilon)$, where $\eta^{-1}(.)$ describes how reported values differ from what respondents believe the actual measure is. Although ϵ is not correlated with y, it could be correlated with b.

[11] This is akin to the assumptions made in the principal-agent literature to allow the use of the first-order approach to solving maximization programs (see Milgrom, 1981, and Rogerson, 1985). A sufficient, but not necessary, condition for this assumption is the first-order stochastic dominance of $F(b|\tilde{b}'')$ over $F(b|\tilde{b}')$.

[12] We test A1 for each race and gender group in NHANES using nonparametric regressions of actual height and weight on their reported values. We find no statistically significant violations of monotonicity in any case.

Now consider the unconditional distribution of true values, $F(b)$.[13] Both $F(b)$ and $G_j(\tilde{b}_j)$ are continuous, monotonically increasing functions with ranges in the interval $[0, 1]$. This implies that for every value of $\tilde{b}_j$ there is a b such that

$$F(b) = G_j(\tilde{b}_j).$$

In general, $b \neq \tilde{b}_j$; however, taking the inverse of $F(\cdot)$, we have:

$$b = F^{-1}(G_j(\tilde{b}_j)),$$

which maps reported values into the true values of b.

Note that $F^{-1}(\cdot)$ does not depend on whether the reported values, $\tilde{b}_j$, are drawn from the primary or auxiliary sample. It simply takes the percentile rank associated with a reported value in sample j and returns the actual value that has the same position in the distribution of true values. As a result, $F(b|G_V(\tilde{b}_V)) = F(b|G_P(\tilde{b}_P))$. This implies that the percentile ranks, $G_j(\tilde{b}_j)$, satisfy transportability, even when the reported values do not. Therefore, the percentile rank approach satisfies both of the conditions required for the use of validation data.

3   Data and the Transportability of Self-Reported Measures

This section begins with a brief introduction to the data sets we use in our analysis. We then demonstrate that self-reported height and weight do not satisfy transportability between these data sets.

---

[13] $F(\cdot)$ does not vary between datasets when the datasets are both representative samples of the same population.

3.1 Three Data Sets

We draw our primary datasets from two sources, the Behavioral Risk Factor Surveillance System (BRFSS) and the American Time Use Survey (ATUS). The BRFSS is a telephone survey conducted by the Centers for Disease Control in conjunction with state health departments. It focuses on health and risky behaviors, but also contains a variety of demographic information. The primary advantage of the BRFSS for obesity studies is its size. With over 300,000 respondents per year in the later waves, the BRFSS is large enough to compute reliable state-level descriptive statistics. Additionally, the large sample size makes it popular among economists seeking to identify causal effects using inherently inefficient estimation techniques such as instrumental variables.

The ATUS was designed to measure how people spend their time rather than to study health outcomes. It is a telephone survey that asks respondents to sequentially report what they did on the day prior to the interview. In addition to the time diary, the ATUS collects demographic information about respondents and members of the respondent's household, and employment status information for the respondent and the respondent's spouse. In 2006, 2007, and 2008, the U.S. Department of Agriculture's Economic Research Service (ERS) sponsored Eating and Health Modules that collected information about the respondent's health, including weight and height, and additional information on time spent eating and drinking.

Following the previous literature, our external validation data are drawn from the National Health and Nutrition Examination Survey (NHANES). NHANES is collected by the Centers for Disease Control and Prevention with the goal of assessing the health status and behaviors of children and adults in the United States. While BRFSS and ATUS only include self-reported measures of height and weight, respondents in NHANES were measured and weighed

after they were asked to report height and weight. We use data on race, gender, and age from the demographic background files of the 2007-2008 wave; reported values of height and weight from the Weight History questionnaire; and measures of actual height and weight from the physical examination.

For the sake of consistency, we restrict the samples from all three data sets to 2007 and 2008. We also focus on respondents between the ages of 19 and 64 who identify as Caucasian, African-American or any other racial or ethnic group.[14] All estimation uses weights to ensure that each sample is representative of the same population.

Table 1 presents basic summary statistics for the three datasets. Even with all race and gender groups pooled together, average reported weight is higher in the NHANES (180 lbs) sample than either BRFSS (178.6) or ATUS (178). Average actual weight, on the other hand, is roughly two pounds heavier for the full NHANES sample.

The demographic variables in Table 1 are very similar across samples, which is consistent with the samples being representative of the same populations. The one notable exception is that black respondents appear to be underrepresented in the BRFSS. Such differences may point to shortcomings in the sample weights used; however, any shortcomings in our sample weights would affect both of the correction methods we compare.[15] Therefore, we ignore this difference in the work that follows.

---

[14] All respondents who identify as Hispanic are included in the "other" category. We did not divide this category further due to the sample size in NHANES. For the sake of convenience, we refer to these groups as though they are defined by race, even though that is not strictly correct.

[15] Additionally, our point can also be demonstrated using only data from race and gender groups that appear to be equally represented across samples.

3.2 The Transportability of Self-Reported Height and Weight

The most relevant difference between the samples is that respondents in the BRFSS and ATUS are surveyed by telephone, while respondents in the NHANES are interviewed in person with reason to believe they will be weighed and measured.[16] It seems natural to expect misreporting to be more severe in telephone surveys than in surveys conducted in person prior to a medical examination. Pinkston (2013) points out that phone interviews in the NLSY cohorts are associated with lower reported weights for white women than in-person interviews, even though NLSY respondents have no reason to believe they will be measured in either case. Any such difference in misreporting, however, implies that $E(b|\tilde{b}_P) \neq E(b|\tilde{b}_V)$ when $\tilde{b}_P = \tilde{b}_V$ and that the transportability condition does not hold when using the standard approach.

Comparing differences in misreporting between our data sets is straightforward. We expect the distribution of actual height and weight to be the same in all three datasets because they are all representative samples from the same population. If the distributions of a reported value are not the same in two samples that have the same distribution of actual measures, the relationships between actual and reported measures are also not the same. Therefore, a simple comparison of reported height and weight across samples is sufficient to evaluate whether the key assumption of the standard approach holds.

Figures 1A compare the densities of reported height and weight across samples for white, black and other women. Relative to white women in NHANES, white women in both BRFSS and ATUS report weights between 120 and 150 pounds more frequently and report higher weights less frequently. The comparison of reported weight produces a similar picture for women of other races. We only see a difference in the reported weights of black women in the

---

[16] See Han et al. (2009) and the online documentation for NHANES.

upper half of their distributions; however, self-reported height appears more sensitive to context for black women than for white or other.

Figures 1B provide analogous comparisons for men. Differences in reported weight are less pronounced for men than for women; but differences in height appear to be larger, especially for black and other men. In sharp contrast to the results for women, black men and men from other races appear more likely to report higher weights when interviewed on the phone than when interviewed in person.

Tables 2A and B compare the self-reported measures in BRFSS and ATUS to those from NHANES. Each table presents averages of reported height and weight, followed by the medians, 75th and 90th percentiles. The final row of each table then presents nonparametric Kolmogorov-Smirnov tests for the equality of distributions between samples.

In both tables, we see that women reported lower weights over the phone than in person, but this result is more consistent for black and white women than for women of other races. The differences in average reported weight appear to driven more by the upper tail for black and (to a lesser degree) other women than for white women.

The differences in reported distributions are less pronounced for men. We only see differences in reported weight for black men and men of other races, and again they appear to report higher weights when they cannot be observed by the interviewer. There are small differences in reported height, but they are less obvious in the summary statistics than the kernel densities.

The Kolmogorov-Smirnov tests reject the equality of at least some distributions for every race and gender group. The only cases in which we cannot reject the equality across samples of

either measure are for white and black men in BRFSS compared to NHANES; however, even then we can reject the equality of reported BMI distributions (not shown) for black men. Overall, these results strongly suggest that the key assumption of the standard approach does not hold, and that NHANES data cannot be reliably used as validation data for the BRFSS and the ATUS when using the standard approach.

Finally, it is worth noting that some differences in the results for BRFSS and ATUS may be expected. Although both are conducted over the telephone, they have very different focuses. The differences in the reported height of black men in ATUS and BRFSS (Figure 1B), for example, could be caused by something as simple as differences in the preceding questions or in the surveys' introductions.

4   Comparing Methods for Predicting BMI

The standard method used by Cawley (2002, 2004) to correct BMI for measurement error involves regressing actual height and weight in NHANES on respondents' reported values, and then predicting the actual measures in the primary dataset. Specifically, for each race and gender category, we regress actual height (or weight) on cubic polynomials in age and reported height (weight). We then predict actual height and weight, and use those values to calculate predicted BMI.[17]

---

[17] L&S argue that constructing a nonlinear function of mismeasured variables from predicted values of those variables may provide a useful approximation, but predicting the nonlinear function directly is preferable. In the case of BMI, this means that researchers who are interested in BMI should predict BMI directly instead of constructing it from predicted height and weight; however, we find that the mean squared error associated with the prediction of BMI is higher in NHANES when BMI is predicted directly than when it is constructed from predicted height and weight. Furthermore, predicting height and weight is more consistent with the previous obesity literature. Therefore, we do not follow the advice of L&S in this particular case.

The estimation used in our method is similar, but uses the percentile rank of the reports and a more flexible functional form. Consistent with the fact that percentile ranks are (roughly) distributed uniformly between zero and one, we found that regressing the actual measures on simple polynomials of the percentile ranks resulted in predicted values in NHANES that were poor fits for the actual measures.[18] For that reason, we regress the actual measures on cubic b-splines in the percentile rank of reported values.[19]

Figures 2A through 2F compare kernel densities of predicted BMI using our percentile-rank method to predictions using the standard validation approach for each race and gender group. Each figure contains four graphs, one for each prediction method and primary data set. Each graph compares predicted BMI from a primary data set to the analogous prediction from NHANES, as well as actual BMI from NHANES. At the bottom of each figure, we include results for Kolmogorov-Smirnov tests of the differences observed in each graph.

Figure 2A shows that the standard method can produce values of predicted BMI that differ significantly between samples. These differences follow the differences between samples in reported weight, and the Kolmogorov-Smirnov tests strongly reject the hypothesis that the standard method produces predicted values in ATUS or BRFSS that are equal to those in NHANES. Again, this suggests that the standard method is inappropriate in our context.

In contrast, the density functions of BMI predicted using the percentile rank method are very similar across samples. The density of predicted BMI in ATUS is almost indistinguishable from the analogous density in NHANES. The Kolmogorov-Smirnov tests both have p-values

---

[18] The distributions of percentile ranks differ from uniform distributions because reported values of height and weight have clustering at certain numbers (e.g., intervals of five pounds).

[19] The details of our estimation are included in the Appendix.

over 0.8, providing no reason to doubt that the density functions of predicted BMI are the same across samples when our method is used.

The results for black women (Figure 2C) and women of other races (2E) are similar to those for white women. The densities of predicted BMI using our approach are noticeably closer to the corresponding densities in NHANES, which makes them more similar to actual BMI in NHANES (and presumably the national population). Kolmogorov-Smirnov tests again reject the equality of the standard method across data sets, but fail to reject our approach.

The results for men are less striking than the results for women, but still support our approach. The density functions for white men (Figure 2B) appear more similar when the percentile method is used than when the standard method is used, but testing the equality of distributions suggests that none of these differences are statistically significant. Although the graphs of kernel densities for black men do not tell an obvious story, the K-S tests reject the standard approach and fail to reject our percentile method.[20] Finally, the standard method is rejected for men of other races when we compare the ATUS to NHANES, and the percentile method again produces predictions of BMI that are more consistent across contexts.

As a robustness check, we also tested for differences in the density functions of predicted BMI between the ATUS and BRFSS (not shown). We would not expect our percentile correction to produce different distributions between the two primary samples, and we find no evidence that it does. On the other hand, we find statistically significant differences for men and women of

---

[20]Regardless of method, the BMI of black men appears to be more difficult to predict that the BMI of other groups. This is appears to be due to errors in the prediction of height for black men.

other races in these surveys when the standard correction is used, which suggests that differences in context that are less obvious than mode of collection may affect misreporting.[21]

5   Do These Differences Matter?

The results of sections 3 and 4 strongly suggest that using the relationship between reported and actual height and weight in the NHANES to predict BMI in data from the ATUS or BRSS is inappropriate. Furthermore, the alternative we propose appears to work well in practice, producing predictions of BMI that are consistent across random samples of the same population. In this section we provide results that demonstrate the potential importance of our adjustment to empirical work.

5.1   Effects of BMI Correction on Regression Estimates

This section is intended to illustrate the impact our adjustment might have on empirical work that uses BMI or obesity in regressions. For both the BRFSS and ATUS, we consider one case where BMI or obesity is an explanatory variable and one case where it is the dependent variable. We report the results both for the full sample and for each race×gender subgroup. None of the estimates presented in this section consider endogeneity or any other complications researchers might encounter. These results are presented for demonstration purposes only.

To assist with the interpretation of results, Table 3 presents average BMI and percentages overweight, obese (BMI≥ 30), and class II/III obese (BMI≥ 35) by correction method and race×gender group for both the BRFSS and ATUS samples. Averages for the actual measures from NHANES are included for comparison.

---

[21] Kolmogorov-Smirnov tests reject the equality of distributions of BMI predicted using the standard correction with *p*-values less than or equal to 0.001 for other women and men. No other test suggests a statistically significant difference in predicted values between any ATUS and BRFSS subsamples.

In the full sample of either dataset, correcting for measurement error increases average body mass, and the effects are most pronounced for the upper tail of the distributions. Roughly 10.5% of population reports a BMI of 35 or higher. Applying the standard correction raises this to 12.5%. Applying our percentile-based correction produces estimates of nearly 14%, which is consistent with the rate of class II/III obesity using measured height and weight from NHANES.

The patterns seen in the full sample are seen to some degree in most of the race and gender groups, but they are most pronounced for white women. The average reported BMI for this group is 26.9. The standard correction increases average BMI to 27.6, and our correction increases it further to 28.2, which is the same as the average actual BMI from NHANES. The differences in the correction methods are larger at the upper tails of the distribution. For example, the overall obesity rate based on uncorrected reports is 25.5%. The standard correction increases the rate to 30%, and our correction increases it further to 32.7%. The actual rate from NHANES is 33.8%. Thus our correction still slightly underestimates the obesity rate for white women, but it does significantly better than the standard approach. Looking at class II/III obesity, the rate is under 11% when self-reported measures are used, just over 13% using the standard correction, and around 16% using our percentile-based correction, which again is lower than the actual rate from NHANES (17.3%).

Black men are the most obvious exception to the pattern seen in the full sample. The average self-reported BMI in both the BRFSS and the ATUS samples is nearly indistinguishable from average measured BMI in NHANES for black men. Furthermore, while black men appear to under-report obesity and class II/III obesity slightly, the incidence of self-reported overweight status is higher than the actual incidence for black men.

5.1.1 BRFSS: Food Prices and BMI

For our first empirical example, we use the BRFSS to evaluate the relationship between state-level food prices and BMI. Table 4 reports coefficient estimates of interest from OLS regressions of BMI on food price and other basic control variables, as well as probit estimates (presented as average marginal effects) of the effects of food prices on the probabilities of being obese or class II/III obese. In each case, results are presented with no adjustment, adjustment using the standard approach, and then adjustment using our percentile rank method. The control variables include race (dummies for non-Hispanic black and non-Hispanic white), education (dummies for some high school, high school, some college, and four-year college degree or greater), marital status (dummies for married, divorced, and widowed), age, inflation-adjusted household income, and a dummy for whether the year was 2008. The state food price measure is computed from city-level data from the Council for Community and Economic Research (formerly American Chamber of Commerce Researchers Association) Cost of Living Index.[22] The sample average food price is $2.56, so a one-unit increase in food price represents approximately a 40% increase relative to the mean.

The results presented in Table 4 suggest that neither the standard correction nor our correction affect the conclusion that higher food prices are associated with lower body mass; however, the choice of correction method has potentially important implications for the magnitudes of coefficients. For the full sample and for most subgroups, the standard correction leads to larger magnitudes than no correction, while our correction leads to even larger magnitudes than the standard correction. This is consistent with the aforementioned result that

---

[22] Following Chou et al. (2004), for each city we average over the prices of each grocery food item, weighting by the C2ER shares of each item's importance in the basket of goods. We then define state prices as the population-weighted average of the prices in the state's C2ER markets. Prices are in 2008 dollars.

measurement error in self-reported weight and height serves to compress the BMI distribution. The more measurement error is eliminated, the more "stretched out" the BMI distribution becomes. In other words, the smaller the measurement error, the larger the change in BMI that is associated with a given change in food prices. For the 0-1 variables, this stretching out of the distribution increases the number of lower-BMI individuals (i.e. those closer to the cutoff) that are categorized as obese or class II/III obese. It is therefore not surprising that our correction, which purges the most measurement error, would lead to larger magnitudes than the standard correction, which purges some measurement error, which in turn leads to larger magnitudes than using no correction.

More specifically, consider the whole-sample regressions. The effect of food price on BMI using our percentile rank correction is 23% larger (-0.937 compared to -0.764) than that using no correction, and 9% larger (-0.937 compared to -0.858) than that using the standard levels-based correction. For obesity, the estimated effect using our approach is 20% larger than with no correction, and 3.4% larger than with the standard correction. The differences are most striking, however, for class II/III obesity. The estimates with our correction are 64% and 28% larger than those using no correction and the standard correction, respectively. The finding that mitigating measurement error matters most for class II/III obesity makes sense in light of the aforementioned results from the literature that the extent of misreporting of both weight and height increases as weight increases (Rowland, 1990; Cawley, 2002). In other words, correcting measurement error leads to the largest increases in BMI among the right tail of the distribution, where the Class II/III obesity cutoff lies.

Accurately estimating effects on class II/III obesity is vital, as a recent meta-analysis has shown that an increased risk of mortality from high BMI does not begin until crossing the class

II/III obesity threshold (Flegal et al., 2013). To illustrate, suppose we are interested in predicting lives saved from a calorie tax that raises the price of the food basket by $1. Obesity is estimated to cause 112,000 deaths per year (Flegal et al., 2005), and Flegal et al.'s (2013) results suggest it is reasonable to attribute all of the premature mortality from obesity to class II/III obesity. The class II/III obesity rate in the 2007-2008 NHANES is 14.4%. Therefore, using the estimate with our correction of -0.041 a $1 increase in food prices would reduce class II/III obesity by 28.5%, compared to 22.2% using the standard correction's estimate of -0.032 and 17.4% using the estimate with no correction of -0.025. Multiplying these numbers by the annual deaths from obesity, the estimated lives saved from the hypothetical policy are 31,889 using our correction, compared to 24,889 using the standard correction and 19,444 with no correction. Therefore, the chosen correction method can lead to important differences in policy implications, even if the general conclusions about sign and statistical significance are unaffected.

Turning to the subsamples, we observe the same general pattern of the magnitudes increasing as measurement error is purged for all groups except black women, black men, and men of a race other than white or black. Even for these three groups, however, our correction still leads to the largest magnitudes for class II/III obesity – substantially larger magnitudes for black women and other men. For BMI, our correction increases the food price effect most substantially relative to the standard correction for other men (14% larger estimated food price effect), black women (14% larger), and white women (13%). For obesity, we observe the largest percentage increases in magnitudes using our correction as opposed to the standard correction for black women (30%) and white women (11%). Importantly, for class II/III obesity our correction increases the estimated food price effect by over 20% relative to the standard correction for all

groups except men. The largest changes are among other men (44%), white women (41%), and black women (37%).

5.1.2   ATUS: Food Prices and BMI

Table 5 is analogous to Table 4, but presents estimated effects of food prices on body mass using data from the ATUS. Overall, the results are very similar, but the standard errors are larger in Table 5 due to the smaller sample size. Looking at the entire sample, the estimated effect of food prices on BMI is smallest when reported BMI is used and largest when our rank-based correction is used. Furthermore, correcting for measurement error appears to have the largest effect when we consider class II/III obesity. The estimated effect of an increase in food prices on the probability of class II/III obesity is 29% larger with the standard correction than with reported values (-0.036 versus -0.028), and 22% larger with our correction than with the standard correction (-0.044 versus -0.036).

Looking within race and gender groups, the results for white women again show that correcting for measurement error results in larger estimated effects of prices on body mass. The estimated effects of food prices on BMI fall for black men as we correct for measurement error, which is consistent with black men reporting themselves as overweight more often than they are.

5.1.3   BRFSS: BMI and Diabetes

We next turn to an examination of the implications of our percentile rank correction in regressions with a weight-related independent variable. For the BRFSS, we consider a question of broad interest to epidemiologists and health policy researchers: the impact of obesity on diabetes. We estimate probit models with a dummy for whether the individual has ever been diagnosed with diabetes as the dependent variable; either BMI, obese, or class II/III obese as the

independent variable of interest; and the same set of controls as our preceding BRFSS analysis in Section 5.1.1.

Average marginal effects of BMI, obesity, and class II/III obesity on P(Diabetes) are reported in Table 6. In all regressions, the association between BMI, obesity, or class II/III obesity and diabetes is positive and statistically significant, so the correction method again does not influence the general conclusions. However, important differences again emerge in the magnitudes. In most regressions, the standard correction leads to smaller magnitudes than no correction, while our correction leads to even smaller magnitudes than the standard correction. This is the opposite of the pattern observed when the weight-related variable was the outcome, and is again consistent with the observation that measurement error in self-reported weight and height compresses the BMI distribution. Correcting measurement error leads to a larger change in BMI being associated with a given change in diabetes, and therefore a smaller coefficient estimate when BMI is an explanatory variable.

Turning to the more specific results for the full sample, the most interesting observation is that the corrections have only a minimal effect on the estimated relationship between BMI and diabetes, but more substantial effects on the estimates for obesity and class II/III obesity. Being obese is estimated to increase P(Diabetes) by 6.9 percentage points using no correction, 6.5 percentage points using the standard levels correction, and 6.2 percentage points using our percentile-rank correction. Our correction therefore leads to a 10% smaller magnitude than no correction, and a 5% smaller magnitude than the standard correction.

For most subsamples, we observe the same pattern of the estimated effects of obesity and severe obesity on P(diabetes) decreasing with the extent of the measurement error purged. The most notable exceptions are for black men, but they are also the only group for which the

incidence of class II/III obesity appears to be lower when our correction is used than it is when the standard correction is used. In the BMI regressions, our correction makes the biggest difference for black women (10% smaller magnitude than no correction, 8% smaller magnitude than the standard correction). In the obesity and class II/III obesity regressions, our correction is most consequential for white women. The estimated effect of obesity on P(diabetes) using our correction is 14% smaller than using no correction, and 7% smaller than using the standard correction. For class II/III obesity, these numbers are 10% and 6%, respectively. It is not surprising that the correction is important for white women since they are the group among which underreported weight is the most common.

5.1.4   ATUS: BMI and Disability

The results in Table 7 provide another example of how correcting for measurement error can affect estimates that use a measure of body mass as an independent variable. In this case, we use data from the ATUS to consider effects of BMI, obesity and class II/III obesity on the probability of being out of labor force due to disability.[23] The results in Table 7 are average marginal effects from probit models that use the same controls as previous regressions in the section.

As in the case of diabetes, adjusting for measurement error does not affect the basic conclusion that body mass is positively associated with disability, but it does reduce the estimated effects. The average marginal effect of obesity on disability falls by 26%, from 0.024 (0.005) to 0.018 (0.004), when the standard correction is applied; and by 32%, to 0.016 (0.004) when our percentile correction is used. The difference between our correction and the standard correction is more pronounced when we consider the effects of class II/III obesity. The estimated

---

[23] This variable equals one if the respondent reported being out of the labor force due to disability in either the ATUS or the CPS. Its mean (standard deviation) is 0.0559 (0.2297). We acknowledge that this is variable is not an ideal measure of disability; however, it is still useful for the purpose of our demonstration.

effects fall by 22% when our correction is used, but by only 6% when the standard correction is used. The results, therefore, are again consistent with measurement error compressing the upper tail of the BMI distribution and resulting in fewer people being classified as obese or class II/III obese.

The patterns of effects for most race and gender groups are similar to those for the entire sample. The results for white women again mirror those for the full sample most closely; however the results for black women may be the most striking. The average marginal effect of class II/III obesity on the probability a black woman is disabled falls by nearly 26% when the standard correction is applied, but by more than half when our percentile rank correction is used.

5.2 Obesity Maps

Finally, we demonstrate the effect of measurement error correction on the estimated prevalence of obesity in the United States. Figure 3 is inspired by the well-known obesity maps produced by the CDC using data from BFRSS.[24] Figure 4 is similar to Figure 3, but we use the standard validation method to correct for measurement error before calculating the prevalence of obesity. Finally, Figure 5 shows the map after our percentile rank method has been used.

These three figures demonstrate that correcting for measurement error can have dramatic effects on the estimated prevalence of obesity. Our percentile rank correction has a larger effect than the standard validation approach because the standard approach is biased by the measurement error it aims to correct. Looking at the nation as a whole, we find that 27.3% of the population reports being obese, 30.8% are found to be obese using the standard correction, and 33.3% are found to be obese using the correction we propose.

---

[24] There are a couple minor differences between our maps and those produced by the CDC. While the CDC map considers all states and uses data on all adults, we focus on adults between the ages of 19 and 64 in the continental United States. We also pool data from 2007 and 2008 into one map instead of creating separate maps for each year.

Table 8 summarizes the pattern of obesity prevalence in Figures 4-5. When no correction is made, the obesity rates for most states (34) are higher than 25 percent and less than or equal to 30 percent, while no state has an obesity rate over 35 percent. With the standard correction, most states (29) fall into the 30-35 percent range and three fall into the 35-40 percent range. Using our correction, the modal obesity rate still falls into the 30-35 percent range, but now 15 states have obesity rates in the 35-40 percent range. Furthermore, we find at least a quarter of the adult population in every state is obese when our correction is applied.

It is also interesting to look at the number of states that move to a higher interval of obesity prevalence when we correct for measurement error. When using the standard correction, 36 states move to the next highest prevalence interval, while 12 states do not change ranges. Using our approach, no state remains in the same prevalence interval and 14 move up two intervals, which is consistent with a median increase in the prevalence of obesity (not shown) of just over six percentage points.

Effects of correcting for measurement error also vary by state. The largest absolute increases in prevalence when our correction is applied are in West Virginia and Iowa (7.5 and 7.1 percentage points, respectively). The smallest increases, 5.0 and 5.2 points, are in New York and Rhode Island. As a percent of reported obesity prevalence, estimated prevalence increased the most (by over 27 percent) in Delaware and Utah, and the least (around 17.4 percent) in Mississippi and South Carolina. Finally, the corrections increase the variance between states (not shown) from 5.88 to 6.21 with the standard correction and 6.53 with our correction.

6 Conclusion

Since Cawley (2002, 2004) it has been common in the economics-of-obesity literature to correct for measurement error in self-reported height and weight by using data from the NHANES as an external validation sample for the authors' primary data. The standard approach is to regress actual height and weight on reported height and weight using NHANES data, and use the estimated coefficients to predict height and weight in the primary dataset. This approach relies on the assumption that the misreporting of height and weight is the same in both surveys, even though one survey is followed by a physical examination and (in most cases) the other survey interviews respondents by telephone.

We propose an alternative approach that requires much weaker assumptions about the data and is robust to differences in measurement error between samples. Specifically, we assume that if person A says she weighs more than an otherwise similar person B, the conditional expectation of A's actual weight is higher than B's. We show that, if this assumption holds, the relationship between the percentile rank of a respondent's reported height and weight and her actual height and weight will be the same in both samples, even if misreporting is more severe in one sample than the other. If this assumption does not hold, then it is not clear what we can learn from self-reported height and weight.

Our approach replaces the reported measures used in the standard approach, which are surrogates for the actual measures, with surrogates that are more likely to be transportable across datasets. We use the percentile ranks of reported values, as opposed to the reported values themselves, to predict actual height and weight. Our regressions use functional forms of the percentile ranks that are more flexible than the polynomials that are traditionally used with self-reported values, but our approach is otherwise similar to the standard approach. The result is a

correction for measurement error that is more robust than the standard approach, while still being easy to implement.

To illustrate the value of our correction, we compare data from the NHANES and two nationally representative telephone surveys, the BRFSS and the ATUS. First, we compare reported values of height and weight within race and gender subsamples of each dataset to demonstrate that misreporting does vary by context. We find that misreporting appears to be more sensitive to context among groups (e.g., white women) that are more prone to misreporting. When we compare predictions of BMI using the standard approach and our percentile-rank approach, we find that the standard approach predicts statistically significant differences in the distribution of BMI across samples that are representative of the same populations, and those differences reflect the differences in reported values. We find no evidence of such differences between samples when BMI is predicted using our alternative method. In other words, our results confirm that the standard approach is biased by differences in misreporting between samples, while our approach is not.

We also consider how corrections for misreported height and weight might matter in future empirical work. We find that the estimated prevalence of obesity and class II/III obesity is higher when our rank-based correction is used than when either the standard correction or no correction is used. Next, we present examples of regression estimates with body mass as a dependent and an independent variable. In each of these examples, we find differences in coefficient estimates that are consistent with measurement error compressing the upper tail of the distribution. The differences in coefficient estimates between our correction and the standard correction are often similar in size to the differences in estimates between the standard correction and uncorrected

28

reports. These differences are often economically significant and could affect the conclusions researchers and policymakers draw.

There are a few caveats we wish to discuss. First of all, we must acknowledge that an internal validation sample, drawn at random from the primary data, should be preferred to an external validation sample, even when our correction is used. The rank-based correction we propose allows the use of an external validation sample under weaker assumptions than the standard approach does, but it still requires more assumptions than an internal validation sample would. In many cases, however, an internal validation study is not feasible.

More broadly, our work should be seen as a warning (or reminder) that the misreporting of height and weight is sensitive to the context of the survey. The differences in context between the NHANES and the BRFSS or NLSY cohorts, for example, are relatively obvious but still sometimes forgotten. It is also possible that context varies in ways that affect misreporting even within a given survey. For example, there is no reason to assume that misreporting in a survey that spans decades, such as BRFSS, has remained constant over time because social norms may have changed as waistlines expanded. Additionally, some surveys, such as the NLSY, mix telephone interviews with in-person interviews, which can cause differences in misreporting between interviews with the same respondent.[25]

As a final note, we must emphasize that body mass is not the same as body composition. As discussed by Burkhauser and Cawley (2008) and others, body fat percentage or other measures of adiposity may be preferable to BMI in many applications. Unfortunately, measures of body composition are rarely available in large datasets because, like actual weight and height,

---

[25] As mentioned above, Pinkston (2013) notes that white women in the NLSY97 claim to weigh less in years when they are interviewed by phone than in years when they are interviewed in person.

they are more expensive to collect than self-reported weight and height. Therefore, the

widespread use of self-reported BMI is likely to continue in the future, and measurement error

will continue to be a problem as long as self-reported BMI is used.

References

Baum, Charles. "The effects of cigarette costs on BMI and obesity," *Health Economics* 18, no. 1 (2009): 3-19.

Baum, Charles. "Effects of food stamps on obesity," *Southern Economic Journal* 77, no. 3 (2011): 623-651.

Baum, Charles and Christopher Ruhm. "Age, socioeconomic status and obesity growth," *Journal of Health Economics* 28, no. 3 (2009): 635-648.

Bound, John, Charles Brown, and Nancy Mathiowetz. "Measurement error in survey data." *Handbook of econometrics* 5 (2001): 3705-3843.

Burkhauser, Richard and John Cawley. "Beyond BMI: The value of more accurate measures of fatness and obesity in social science research." *Journal of Health Economics* 27, no. 2 (2008): 519-529.

Burkhauser, Richard, John Cawley, and Maximilian Schmeiser. "The timing of the rise in U.S. obesity varies with measure of fatness." *Economics and Human Biology* 7, no. 3 (2009): 307-318.

Carroll, Raymond J, David Ruppert, Leonard A Stefanski, Ciprian M Crainiceanu. *Measurement Error in Nonlinear Models: a Modern Perspective,* 2nd edition. (2006) Chapman & Hall, Boca Raton, FL.

Cawley, John. "Addiction and the consumption of calories: Implications for obesity." Unpublished manuscript (2002), Cornell University.

Cawley, John. "The impact of obesity on wages." *Journal of Human Resources* 39, no. 2 (2004): 451-474.

Cawley, John and Sheldon Danzinger. "Morbid obesity and the transition from welfare to work." *Journal of Policy Analysis and Management* 24, no. 4 (2005): 727-743.

Cawley, John, John Moran, and Kosali Simon. "The impact of income on the weight of elderly Americans." *Health Economics* 19, no. 8 (2010): 979-993.

Chou, Shin-Yi, Michael Grossman, and Henry Saffer. "An economic analysis of adult obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics* 23, no. 3 (2004): 565-587.

Courtemanche, Charles, Garth Heutel, and Patrick McAlvanah. "Impatience, incentives, and obesity." *Economic Journal* (forthcoming).

Eid, J., H. Overman, D. Puga, and M. Turner. "Fat city: Questioning the relationship between urban sprawl and obesity." *Journal of Urban Economics* 63, no. 2 (2008), 385-404.

Fan, Maoyang. "Do food stamps contribute to obesity in low-income women? Evidence from the National Longitudinal Survey of Youth 1979," *American Journal of Agricultural Economics* 92, no. 4 (2010), 1165-1180.

Flegal, Katherine, M. Carroll, R. Kuczmarski, C. Johnson. "Overweight and obesity in the United States: prevalence and trends, 1960-1994," *International Journal of Obesity Related Metabolic Disorders* 22, no. 1 (1998): 39-47.

Flegal, Katherine, Barry Graubard, David Williamson, and Mitchell Gail. "Excess deaths associated with underweight, overweight, and obesity," *Journal of the American Medical Association* 293, no. 15 (2005): 1861-1867.

Flegal, Katherine, Brian Kit, Heather Orpana, and Barry Graubard. "Association of All-Cause Mortality with Overweight and Obesity using Standard Body Mass Index Categories: A Systematic Review and Meta-Analysis." *Journal of the American Medical Association* 309, no. 1 (2013): 71-82.

Goldman, Dana, Darius Lakdawalla, and Yuhui Zheng. "Food prices and the dynamics of body weight" in *Economic Aspects of Obesity*, Grossman, Michael and Naci Mocan eds., Chicago, IL (2011): University of Chicago Press.

Gregory, Christian. "Wages, BMI, and Age." Unpublished manuscript (2010), United States Department of Agriculture Economic Research Service.

Gregory, Christian and Christopher Ruhm. "Where does the wage penalty bite?" in Economic Aspects of Obesity, Grossman, Michael and Naci Mocan eds., Chicago, IL (2011): University of Chicago Press.

Han, Euna, Edward C. Norton, Sally C. Stearns. "Weight and wages: fat versus lean paychecks." *Health Economics*, 18, no. 5 (2009): 535-548.

Lakdawalla, Darius and Tomas Philipson. "The growth of obesity and technological change: A theoretical and empirical investigation." National Bureau of Economic Research Working Paper No. 8965 (2002).

Lee, Lung-fei, and Jungsywan H. Sepanski. "Estimation of linear and nonlinear errors-in-variables models using validation data." *Journal of the American Statistical Association* 90, no. 429 (1995): 130-140.

Majumder, Md. Alauddin. "Does obesity matter for wages? Evidence from the United States." *Economic Papers* 32, no. 2 (2013): 200-217.

Ogden, C.L. and M.D. Carroll. "Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults: United States, Trends 1960-1962 Through 2007-2008," National Center for Health Statistics report. (2010) http://www.cdc.gov/nchs/data/hestat/obesity_adult_07_08/obesity_adult_07_08.htm.

Pinkston, Joshua C. (2013) "The Dynamic Effects of Obesity on the Wages of Young Workers." University of Louisiville, mimeo.

Plantinga, Andrew and Stephanie Bernell. "The association between urban sprawl and obesity: Is it a two-way street?" *Journal of Regional Science* 47, no. 5 (2007): 857-879.

Rowland, M.L. "Self-reported weight and height." *American Journal of Clinical Nutrition* 52, no. 6 (1990): 1125-1133.

Ruhm, Christopher. "Healthy Living in Hard Times." *Journal of Health Economics* 24, no. 2 (2005): 341-363.

Schmeiser, Maximilian. "Expanding wallets and waistlines: The impact of family income on the BMI of women and men eligible for the earned income tax credit." *Health Economics* 18, no. 11 (2009): 1277-1294.

Appendix: How to Use Our Percentile-Based Correction

Although our percentile-based approach may sound more complicated to implement than the standard approach, we want to assure the reader that the added complication is trivial. Our approach adds two simple steps to the standard approach. Each of those two steps requires one line of code in Stata.

First, we find the percentile rank of the reported measure in the relevant subsample. This is easily done in Stata using the "cumul" command by dataset, race or ethnic group, and gender. Using reported weight as an example, we would have:

bysort *dataset race sex*: cumul *reported_weight* [aw= *samp_wt*], g(*wt_rank*) equal

where *wt_rank* is the newly created percentile rank in the distribution of *reported_weight* for the subsample determined by the indicators *dataset, race* and *sex*.

Actual weight must then be regressed on a flexible function of *wt_rank*, and a polynomial in age. We found that simple polynomials in *wt_rank* were not flexible enough to predict actual weight, and used cubic basis splines in *wt_rank* instead. The second step our method adds to the standard approach generates the splines with the user-written command "bspline".[1] For example,

bspline, xvar(*wt_rank*) p(3) gen(*wt_spline*) knots(0, .05, .1, .25, .5, .75, .9, .95, 1)

where *wt_spline* is the prefix of the generated splines.[2]

From this point on, our approach closely resembles the standard approach. Our splines simply replace the polynomial in reported weight that previous authors have used:

---

[1] The command "bspline" was written by Roger Newson. Documentation and code can be found here: http://econpapers.repec.org/RePEc:boc:bocode:s411701
[2] The number and spacing of knots can be adjusted as needed to improve the fit of the predicted values.

reg *actual_weight* (i.*race*##i.*sex*)##(c.*age*##c.*age*##c.*age* c.*wt_spline**) [aw=*samp_wt*], nocons

Weight is predicted in both the validation sample, which contains *actual_weight*, and the primary

sample. Finally, the process can be repeated for height, allowing a predicted BMI measure to be

constructed.[3]

---

[3] Alternatively, BMI itself could be predicted directly following the same approach.

Table 1. Summary Statistics by Dataset

	BRFSS	ATUS	NHANES
Reported Height	67.37	67.21	67.24
	(4.18)	(4.10)	(4.08)
Reported Weight	178.6	178	180
	(44.67)	(43.60)	(45.15)
Actual Height	….	….	66.9
			(3.88)
Actual Weight	….	….	181.8
			(47.18)
White	0.680	0.689	0.674
	(0.466)	(0.463)	(0.469)
Black	0.103	0.114	0.119
	(0.303)	(0.318)	(0.323)
Other Race/Ethnicity	0.217	0.197	0.207
	(0.412)	(0.397)	(0.405)
Male	0.516	0.513	0.504
	(0.500)	(0.500)	(0.500)
Age	41.29	41.34	41.05
	(12.48)	(12.54)	(12.67)
Observations	17,721	539,072	4,113

Notes: Standard deviations are in parentheses. All samples weighted to be representive of adults in the US between the ages of 19 and 64 in the years 2007 and 2008. "Other Race/Ethnicity" includes all respondents who indentify as Hispanic.

Table 2A. Comparison of Self-Reported Values: BRFSS vs. NHANES

		Mean	Percentiles of Distribution			Kolmogorov-Smirnov
			Median	75th	90th	Test (p-value)
			White Women			
Weight	NHANES	163.907	155	185	220	0.003
	BRFSS	159.947	150	180	210	
Height	NHANES	64.828	65	67	68	0.972
	BRFSS	64.857	65	67	68	
			White Men			
Weight	NHANES	199.387	195	220	250	0.591
	BRFSS	200.060	195	220	250	
Height	NHANES	70.690	71	73	74	0.559
	BRFSS	70.731	71	72	74	
			Black Women			
Weight	NHANES	182.627	170	214	250	0.047
	BRFSS	178.571	170	200	240	
Height	NHANES	64.729	65	67	68	0.572
	BRFSS	64.793	65	67	68	
			Black Men			
Weight	NHANES	198.628	190	225	250	0.179
	BRFSS	201.797	195	225	260	
Height	NHANES	70.221	70	72	74	0.317
	BRFSS	70.279	70	72	74	
			Other Women			
Weight	NHANES	154.718	148	175	205	0.498
	BRFSS	153.868	148	172	200	
Height	NHANES	62.893	63	65	66	0.025
	BRFSS	63.171	63	65	67	
			Other Men			
Weight	NHANES	180.930	175	200	230	0.017
	BRFSS	183.433	179	200	230	
Height	NHANES	67.585	67	70	72	<0.001
	BRFSS	68.142	68	71	72	

Notes: All samples weighted to be representative of adults in the US between the ages of 19 and 64 in the years 2007 and 2008. "Other" includes all respondents who identify as Hispanic.

Table 2B. Comparison of Self-Reported Values: ATUS vs. NHANES

| | | Mean | Percentiles of Distribution | | | Kolmogorov-Smirnov |
			Median	75th	90th	Test (p-value)
			White Women			
Weight	NHANES	163.907	155	185	220	0.002
	ATUS	159.348	150	180	210	
Height	NHANES	64.828	65	67	68	0.993
	ATUS	64.773	65	67	68	
			White Men			
Weight	NHANES	199.387	195	220	250	0.270
	ATUS	197.917	190	220	250	
Height	NHANES	70.690	71	73	74	0.020
	ATUS	70.487	70	72	74	
			Black Women			
Weight	NHANES	182.627	170	214	250	0.017
	ATUS	176.208	170	200	237	
Height	NHANES	64.729	65	67	68	0.116
	ATUS	64.434	64	66	68	
			Black Men			
Weight	NHANES	198.628	190	225	250	0.331
	ATUS	198.814	195	225	257	
Height	NHANES	70.221	70	72	74	0.060
	ATUS	69.827	70	72	74	
			Other Women			
Weight	NHANES	154.718	148	175	205	0.001
	ATUS	150.478	140	170	197	
Height	NHANES	62.893	63	65	66	0.764
	ATUS	62.853	63	65	66	
			Other Men			
Weight	NHANES	180.930	175	200	230	0.010
	ATUS	184.299	180	200	238	
Height	NHANES	67.585	67	70	72	0.036
	ATUS	67.931	68	70	72	

Notes: All samples weighted to be representative of adults in the US between the ages of 19 and 64 in the years 2007 and 2008. "Other" includes all respondents who identify as Hispanic.

Table 3. Average Body Mass by Correction Method and Race/Gender Group from BRFSS and ATUS, Compared to Actual Measures from NHANES

	BMI			Overweight			Obese			Class II/III Obese		
	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method
Full Sample		*28.51*			*67.0%*			*33.6%*			*14.4%*	
BRFSS	27.68	28.21	28.49	63.4%	65.9%	67.2%	28.0%	31.4%	33.0%	10.5%	12.5%	13.8%
ATUS	27.71	28.22	28.52	64.0%	66.2%	67.4%	28.5%	31.7%	33.0%	10.7%	12.5%	13.9%
White Women		*28.20*			*60.6%*			*33.8%*			*17.3%*	
BRFSS	26.91	27.63	28.22	53.2%	57.9%	60.8%	25.5%	30.0%	32.7%	10.8%	13.3%	15.8%
ATUS	26.94	27.65	28.20	53.6%	58.0%	61.0%	25.6%	30.2%	32.6%	10.9%	13.2%	16.0%
White Men		*28.42*			*71.2%*			*31.3%*			*10.2%*	
BRFSS	28.02	28.41	28.45	71.2%	71.8%	72.0%	27.9%	30.8%	31.8%	9.1%	10.9%	11.0%
ATUS	27.98	28.35	28.47	71.8%	72.2%	72.0%	28.2%	30.7%	31.8%	9.0%	10.8%	10.9%
Black Women		*31.50*			*76.5%*			*48.1%*			*28.7%*	
BRFSS	30.13	30.96	31.52	74.0%	76.6%	78.9%	42.2%	46.0%	47.6%	22.0%	24.9%	28.2%
ATUS	30.11	30.90	31.52	74.5%	76.9%	79.5%	42.5%	45.8%	47.8%	21.0%	23.1%	28.2%
Black Men		*28.51*			*65.8%*			*35.9%*			*14.0%*	
BRFSS	28.45	28.60	28.47	69.7%	68.4%	67.0%	33.5%	35.5%	36.1%	12.0%	14.6%	13.6%
ATUS	28.50	28.64	28.48	70.0%	68.8%	66.5%	33.7%	36.1%	35.1%	12.8%	14.6%	13.5%
Other Women		*28.39*			*64.1%*			*35.1%*			*15.7%*	
BRFSS	27.22	27.87	28.25	56.2%	60.8%	63.5%	27.5%	30.7%	32.7%	10.4%	12.1%	14.7%
ATUS	27.06	27.71	28.27	54.0%	57.8%	62.9%	27.3%	30.1%	31.8%	10.3%	11.9%	14.4%
Other Men		*28.13*			*70.8%*			*29.1%*			*9.6%*	
BRFSS	27.75	28.07	28.14	68.4%	69.5%	68.9%	26.5%	29.1%	29.7%	7.9%	8.8%	9.0%
ATUS	27.99	28.30	28.18	70.4%	71.8%	69.4%	28.6%	30.9%	30.2%	8.5%	9.2%	9.1%

Notes: **Actual measures from NHANES are underlined and in italics. All samples weighted to be representive of adults in the US between the ages of 19 and 64.**

Table 4. BRFSS Estimated Effects of Food Prices on BMI, P(Obese), and P(Class II/III Obese) by Race-Gender Group and BMI Correction Method

	BMI			Obese			Class II/III Obese		
	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method
Whole Sample	-0.764 (0.170)***	-0.858 (0.187)***	-0.937 (0.189)***	-0.050 (0.010)***	-0.058 (0.011)***	-0.060 (0.012)***	-0.025 (0.007)***	-0.032 (0.009)***	-0.041 (0.010)***
White Women	-0.704 (0.216)***	-0.786 (0.218)***	-0.890 (0.226)***	-0.042 (0.012)***	-0.046 (0.012)***	-0.051 (0.013)***	-0.021 (0.008)***	-0.027 (0.008)***	-0.038 (0.009)***
White Men	-0.415 (0.191)**	-0.479 (0.193)**	-0.510 (0.208)**	-0.032 (0.015)**	-0.037 (0.015)**	-0.038 (0.015)***	-0.013 (0.011)	-0.014 (0.010)	-0.017 (0.011)
Black Women	-0.894 (0.368)**	-0.871 (0.362)**	-0.991 (0.401)**	-0.029 (0.029)	-0.020 (0.033)	-0.026 (0.029)	-0.050 (0.017)***	-0.049 (0.018)***	-0.067 (0.021)***
Black Men	-1.565 (0.229)***	-1.785 (0.258)***	-1.721 (0.276)***	-0.130 (0.019)***	-0.129 (0.022)***	-0.105 (0.019)***	-0.055 (0.011)***	-0.077 (0.014)***	-0.081 (0.011)***
Other Women	-1.478 (0.321)***	-1.539 (0.301)***	-1.640 (0.334)***	-0.098 (0.022)***	-0.113 (0.026)***	-0.119 (0.025)***	-0.037 (0.010)***	-0.047 (0.013)***	-0.061 (0.013)***
Other Men	-0.710 (0.286)**	-0.855 (0.355)**	-0.975 (0.364)***	-0.042 (0.017)**	-0.063 (0.021)***	-0.061 (0.019)***	-0.028 (0.011)**	-0.036 (0.014)**	-0.052 (0.020)**

Notes: All cells report estimated effects of $1 increase in state food price basket in the corresponding regression; the average food price is $2.56. Average marginal effects are reported in the probit regressions for obesity and Class II/III obesity. Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at 1% level; ** 5% level; * 10% level. The regressions include control variables for race, education, marital status, age, inflation-adjusted household income, and a dummy for whether the year was 2008. The BRFSS sampling weights are used.

Table 5. ATUS Estimated Effects of Food Prices on BMI, P(Obese), and P(Class II/III Obese) by Race-Gender Group and BMI Correction Method

	BMI			Obesity			Class II/III Obesity		
	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method
Whole Sample	-1.099***	-1.179***	-1.251***	-0.083***	-0.083***	-0.084***	-0.028***	-0.036***	-0.044***
	(0.182)	(0.192)	(0.207)	(0.015)	(0.018)	(0.016)	(0.008)	(0.007)	(0.011)
White Women	-1.050***	-1.137***	-1.139***	-0.057**	-0.073***	-0.075***	-0.030	-0.034**	-0.044**
	(0.376)	(0.391)	(0.408)	(0.024)	(0.024)	(0.026)	(0.018)	(0.016)	(0.021)
White Men	-0.605	-0.695	-0.822*	-0.080**	-0.085**	-0.074*	-0.020*	-0.021	-0.008
	(0.430)	(0.458)	(0.487)	(0.036)	(0.043)	(0.041)	(0.012)	(0.016)	(0.014)
Black Women	-1.499	-1.675	-1.800	-0.077	-0.073	-0.084	-0.007	-0.055	-0.097
	(1.079)	(1.169)	(1.294)	(0.071)	(0.082)	(0.088)	(0.052)	(0.046)	(0.067)
Black Men	-1.316**	-1.290**	-1.198**	-0.029	-0.058	-0.028	-0.049*	-0.038	-0.059**
	(0.521)	(0.558)	(0.536)	(0.067)	(0.065)	(0.040)	(0.027)	(0.039)	(0.029)
Other Women	-0.929	-0.901	-0.978	-0.099**	-0.047	-0.055	-0.015	-0.019	-0.055***
	(0.658)	(0.671)	(0.677)	(0.040)	(0.057)	(0.064)	(0.014)	(0.017)	(0.021)
Other Men	-1.876***	-2.051***	-2.174***	-0.130***	-0.119**	-0.150***	-0.047	-0.075**	-0.079**
	(0.692)	(0.690)	(0.683)	(0.049)	(0.049)	(0.045)	(0.034)	(0.035)	(0.032)

Notes: All cells report estimated effects of $1 increase in state food price basket in the corresponding regression; the average food price is $2.56. Average marginal effects are reported in the probit regressions for obesity and Class II/III obesity. Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at 1% level; ** 5% level; * 10% level. The regressions include control variables for race, education, marital status, age, inflation-adjusted household income, and a dummy for whether the year was 2008. The ATUS sampling weights are used.

Table 6 – BRFSS Estimated Effects of BMI, Obesity, and Class II/III obesity on P(Diabetes) by Race-Gender Group and BMI Correction Method

	BMI			Obese			Class II/III Obese		
	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method
Whole Sample	0.0042 (0.0001)***	0.0041 (0.0001)***	0.0041 (0.0001)***	0.069 (0.002)***	0.065 (0.003)***	0.062 (0.002)***	0.107 (0.005)***	0.102 (0.005)***	0.096 (0.004)***
White Women	0.0036 (0.0001)***	0.0035 (0.0001)***	0.0034 (0.0001)***	0.073 (0.002)***	0.068 (0.002)***	0.063 (0.002)***	0.103 (0.003)***	0.098 (0.003)***	0.088 (0.003)***
White Men	0.0044 (0.0002)***	0.0042 (0.0002)***	0.0044 (0.0002)***	0.065 (0.003)***	0.062 (0.002)***	0.061 (0.002)***	0.114 (0.004)***	0.107 (0.004)***	0.105 (0.004)***
Black Women	0.0051 (0.0002)***	0.0050 (0.0002)***	0.0046 (0.0002)***	0.079 (0.004)***	0.077 (0.004)***	0.072 (0.004)***	0.100 (0.007)***	0.098 (0.007)***	0.097 (0.006)***
Black Men	0.0064 (0.0005)***	0.0058 (0.0004)***	0.0060 (0.0004)***	0.081 (0.008)***	0.073 (0.008)***	0.078 (0.009)***	0.125 (0.013)***	0.116 (0.015)***	0.121 (0.013)***
Other Women	0.0040 (0.0002)***	0.0043 (0.0002)***	0.0043 (0.0003)***	0.072 (0.007)***	0.067 (0.006)***	0.068 (0.006)***	0.094 (0.013)***	0.093 (0.012)***	0.093 (0.010)***
Other Men	0.0040 (0.0004)***	0.0041 (0.0004)***	0.0041 (0.0005)***	0.057 (0.005)***	0.052 (0.006)***	0.050 (0.005)***	0.111 (0.017)***	0.099 (0.021)***	0.083 (0.017)***

Notes: All regressions are probits; the cells report average marginal effects of BMI and average effects of a switch from 0 to 1 in obesity and Class II/III obesity status. The dependent variable is a dummy for ever being diagnosed with diabetes; its sample mean is 0.065. Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at 1% level; ** 5% level; * 10% level. The regressions include control variables for race (dummies for non-Hispanic black and non-Hispanic white), education (dummies for some high school but no degree, high school degree but no further, some college but no degree, and four-year college degree or greater), marital status (dummies for married, divorced, and widowed), age, inflation-adjusted household income, and a dummy for whether the year was 2008. The BRFSS sampling weights are used.

Table 7. ATUS Estimated Effects of BMI, Obesity, and Class II/III Obesity on P(Disabled) by Race-Gender Group and BMI Correction Method

	BMI			Obesity			Class II/III Obesity		
	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method	Reported	Standard Method	Percentile Method
Whole Sample	0.0020***	0.0019***	0.0017***	0.0237***	0.0175***	0.0161***	0.0414***	0.0388***	0.0319***
	(0.0003)	(0.0003)	(0.0002)	(0.0045)	(0.0041)	(0.0041)	(0.0042)	(0.0041)	(0.0037)
White Women	0.0027***	0.0026***	0.0025***	0.0369***	0.0295***	0.0239***	0.0533***	0.0498***	0.0442***
	(0.0005)	(0.0005)	(0.0005)	(0.0078)	(0.0076)	(0.0078)	(0.0093)	(0.0090)	(0.0081)
White Men	0.0004	0.0003	0.0004	0.0050	0.0048	0.0032	0.0198**	0.0152*	0.0133*
	(0.0008)	(0.0007)	(0.0006)	(0.0072)	(0.0068)	(0.0067)	(0.0081)	(0.0084)	(0.0080)
Black Women	0.0039***	0.0038***	0.0031***	0.0271	0.0145	0.0164	0.0820***	0.0607***	0.0369**
	(0.0011)	(0.0010)	(0.0009)	(0.0201)	(0.0174)	(0.0152)	(0.0184)	(0.0156)	(0.0148)
Black Men	0.0022	0.0014	0.0013	0.0617*	0.0126	0.0290	0.0709**	0.0781**	0.0622**
	(0.0032)	(0.0028)	(0.0025)	(0.0373)	(0.0290)	(0.0269)	(0.0316)	(0.0394)	(0.0301)
Other Women	0.0026***	0.0026***	0.0026***	0.0345***	0.0350***	0.0345***	0.0411***	0.0413***	0.0337***
	(0.0004)	(0.0005)	(0.0005)	(0.0069)	(0.0067)	(0.0073)	(0.0111)	(0.0105)	(0.0100)
Other Men	0.0020**	0.0023***	0.0022***	0.0229***	0.0198***	0.0223***	0.0431***	0.0430***	0.0373***
	(0.0008)	(0.0008)	(0.0007)	(0.0077)	(0.0072)	(0.0070)	(0.0117)	(0.0099)	(0.0127)

Notes: All regressions are probits; the cells report average marginal effects of BMI and average effects of a switch from 0 to 1 in obesity and Class II/III obesity status. The dependent variable is a dummy for being out of labor market due to disability. Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at 1% level; ** 5% level; * 10% level. The regressions include control variables for race (dummies for non-Hispanic black and non-Hispanic white), education (dummies for some high school but no degree, high school degree but no further, some college but no degree, and four-year college degree or greater), marital status (dummies for married, divorced, and widowed), age, inflation-adjusted household income, and a dummy for whether the year was 2008. The ATUS sampling weights are used.

Table 8: Summary of State-Level Obesity Prevalence Results

	Reported Obesity	Standard Correction	Percentile Correction
Number of States in Each Range of Obesity Prevalence			
20-25%	8	1	0
25-30%	34	15	3
30-35 %	6	29	30
35-40%	0	3	15
Number of States Moving to a Higher Prevalence Category, Relative to Reported Prevalence			
No Change	--	12	0
Up 1 Category	--	36	34
Up 2 Categories	--	0	14

Notes: This table summarizes results presented in Figures 3, 4 and 5. The data are from the 2007 and 2008 BRFSS. Observations are limited to the 48 states in the continental U.S.

Figure 1A. Kernel Densities of Reported Measures for Women
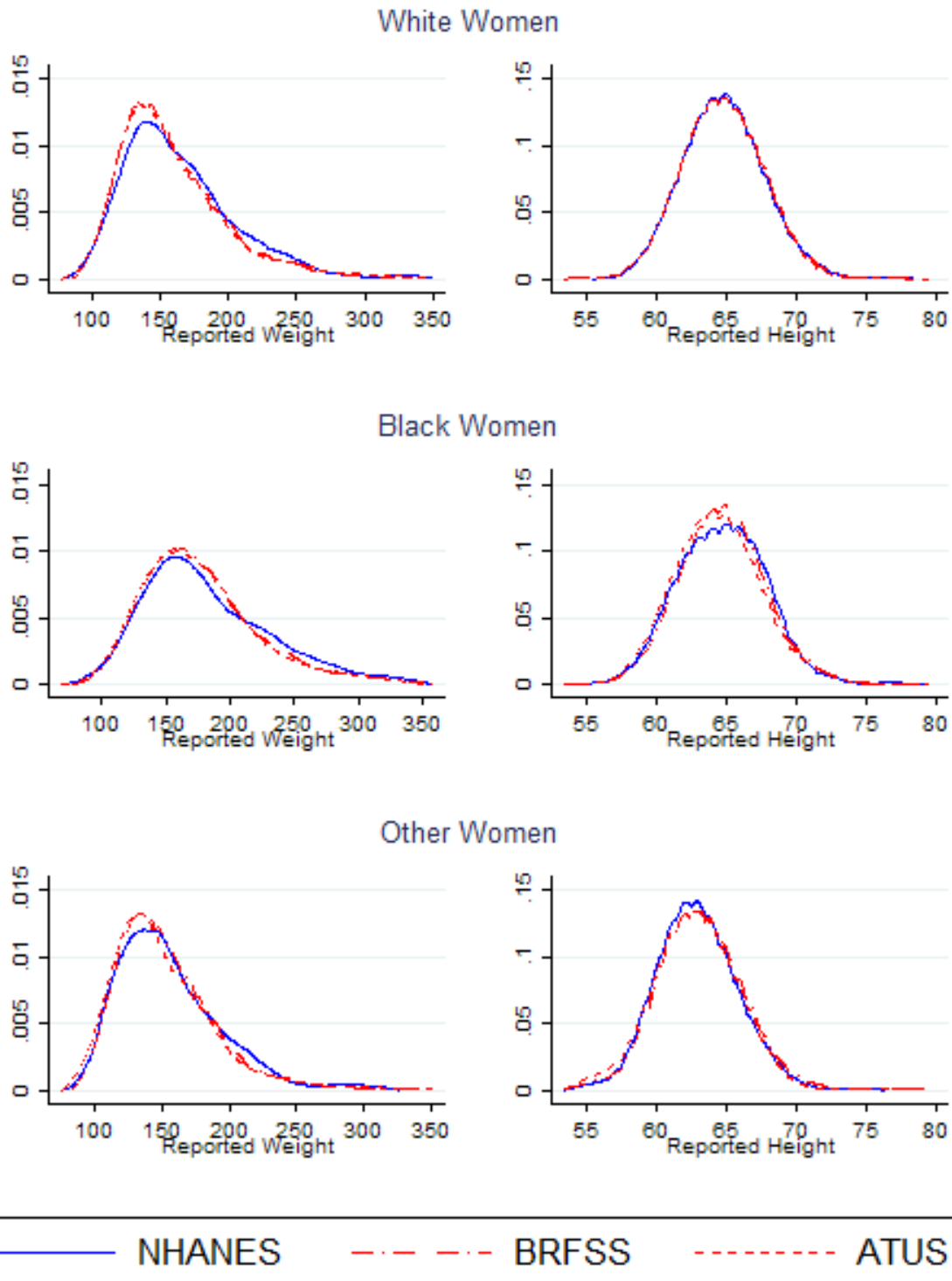
Figure 1B. Kernel Densities of Reported Measures for Men

White Men

Black Men

Other Men
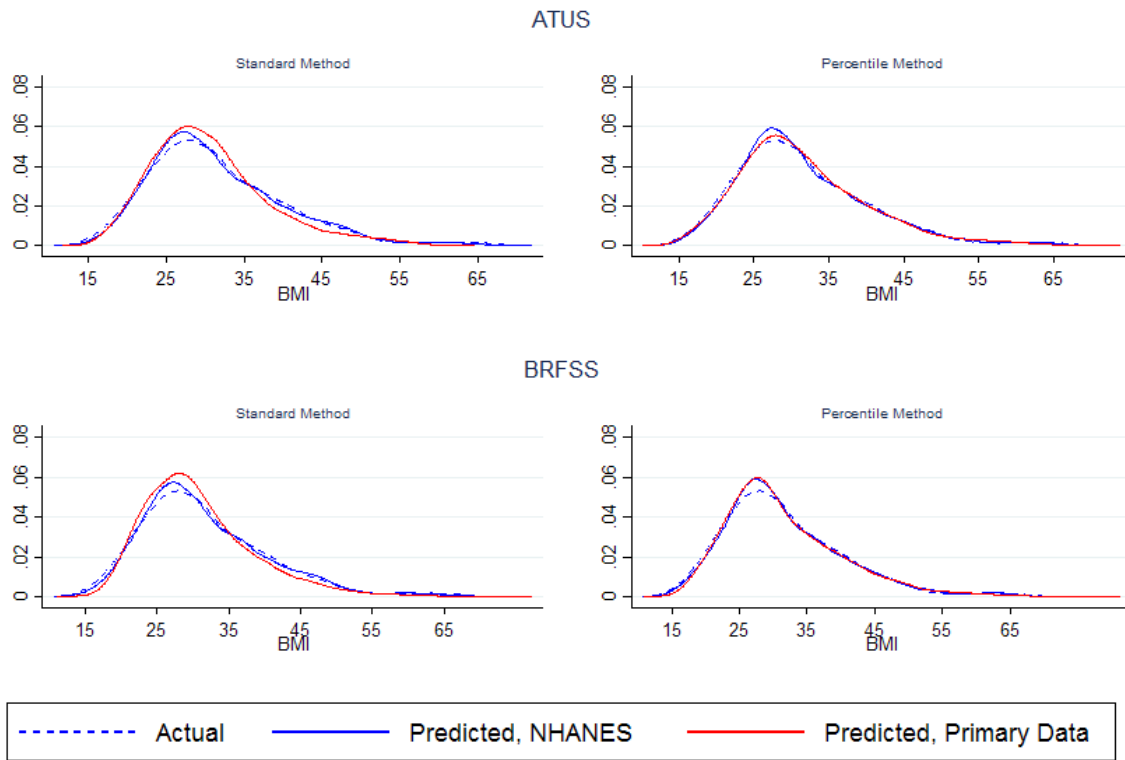

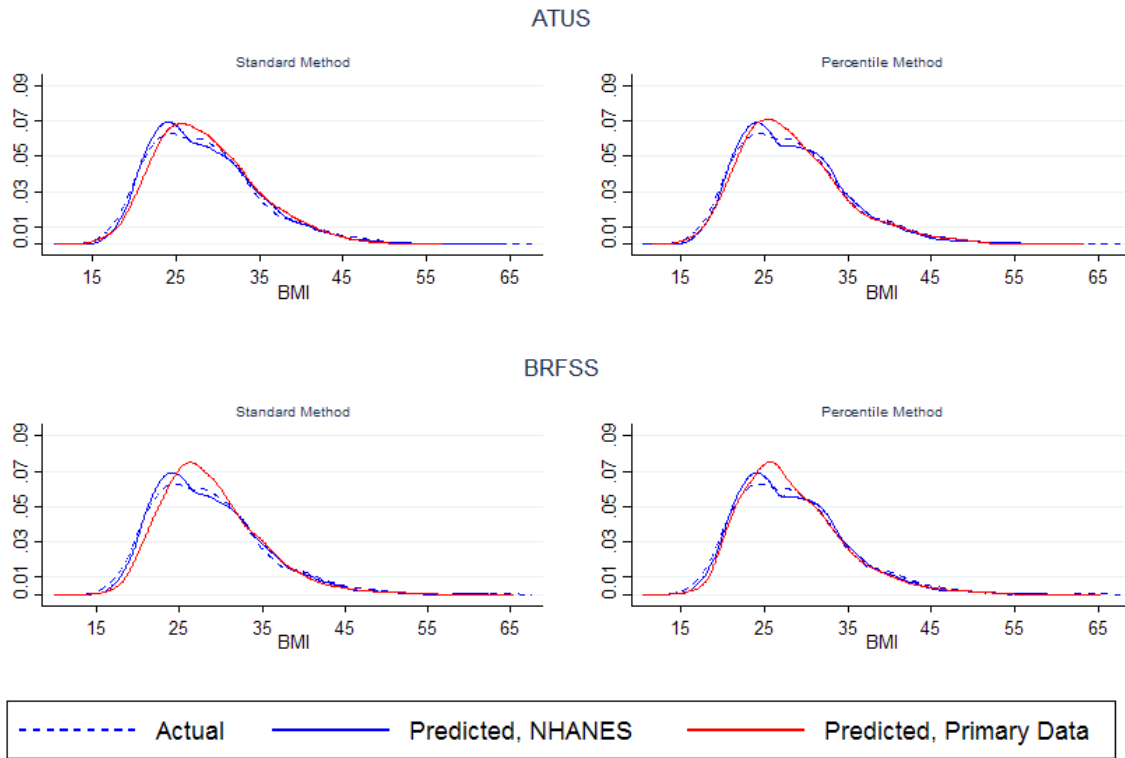NHANES ——— BRFSS —·—·— ATUS ----------
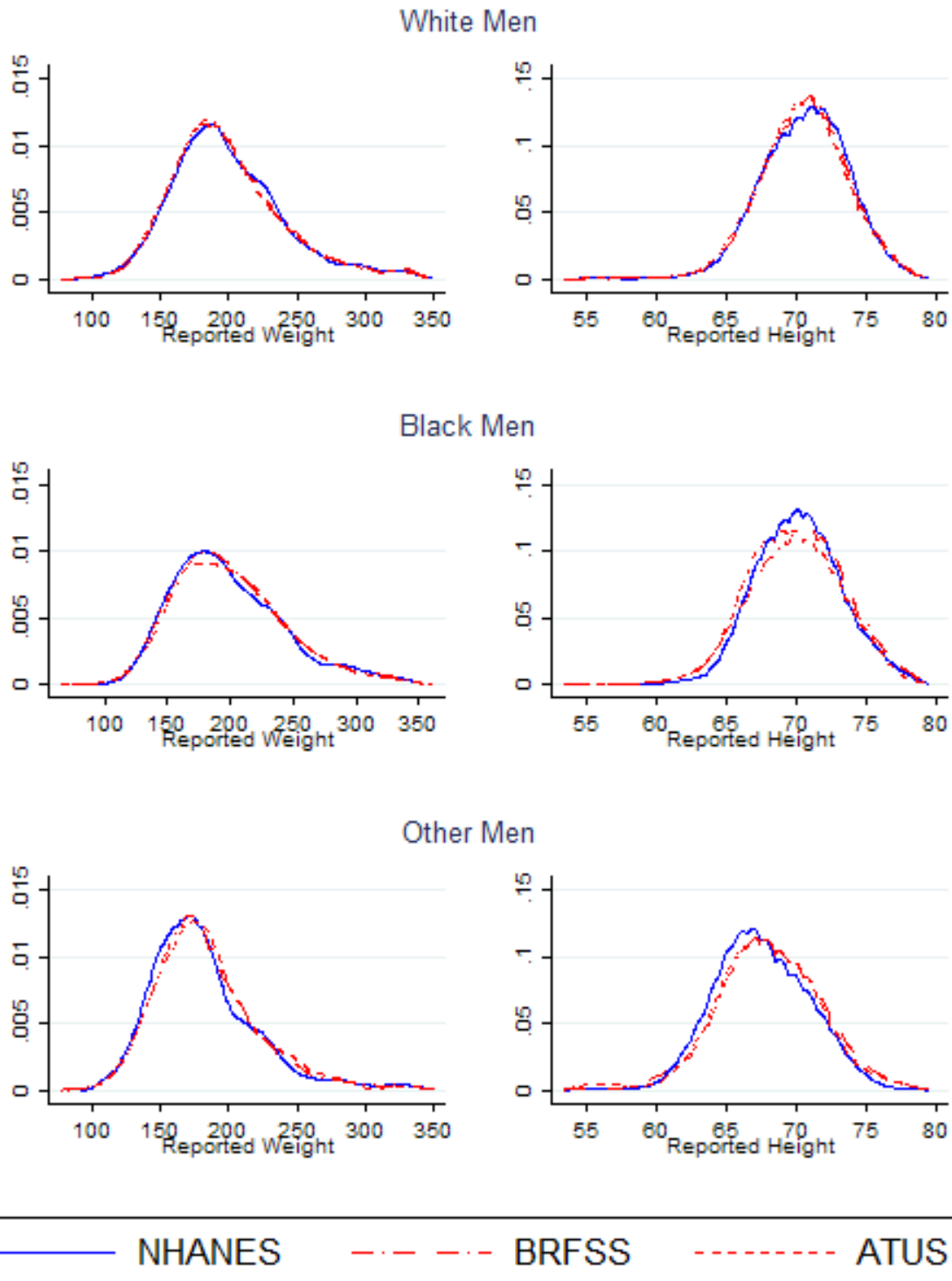
Figure 2A. Standard vs. Percentile Prediction Methods for White Women
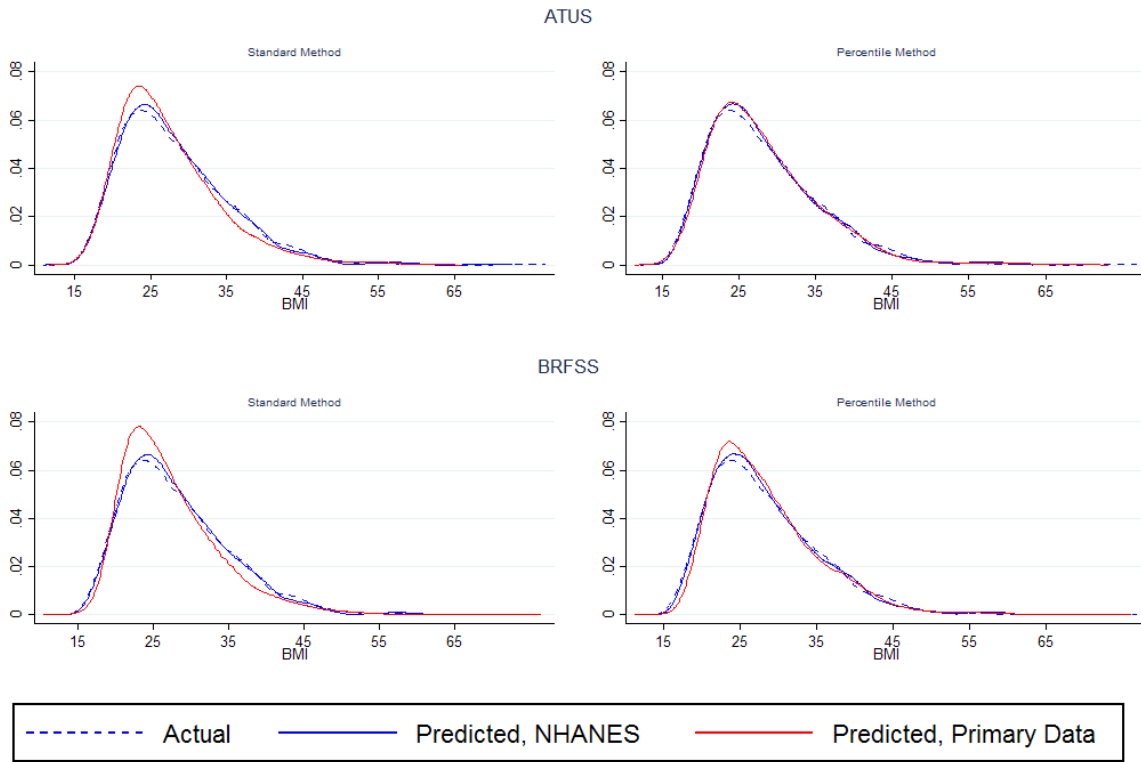
Kolmogorov-Smirnov Tests (*p*-values) for Equality of Density Functions

	Standard Method	Percentile Method
ATUS vs. NHANES	0.006	0.908
BRFSS vs. NHANES	0.006	0.826

Figure 2B. Standard vs. Percentile Prediction Methods for White Men
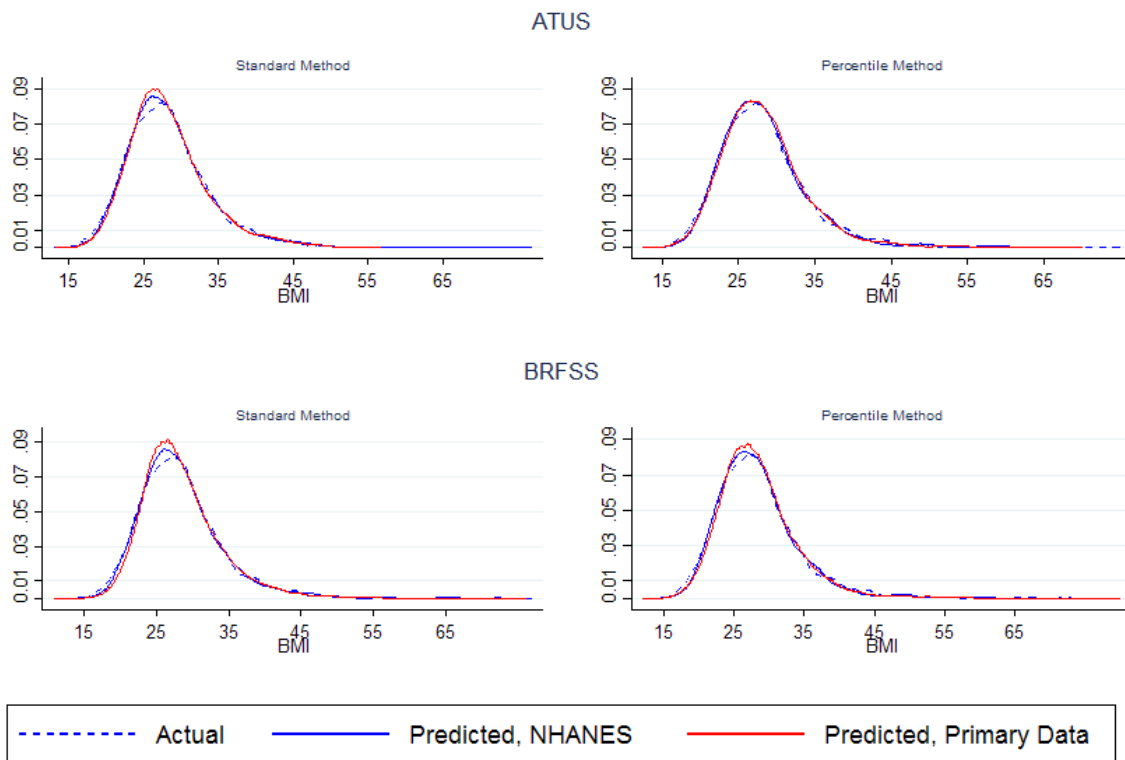
ATUS

Standard Method

Percentile Method

BRFSS

Standard Method

Percentile Method

- - - - - Actual ———— Predicted, NHANES ———— Predicted, Primary Data

Kolmogorov-Smirnov Tests (*p*-values) for Equality of Density Functions

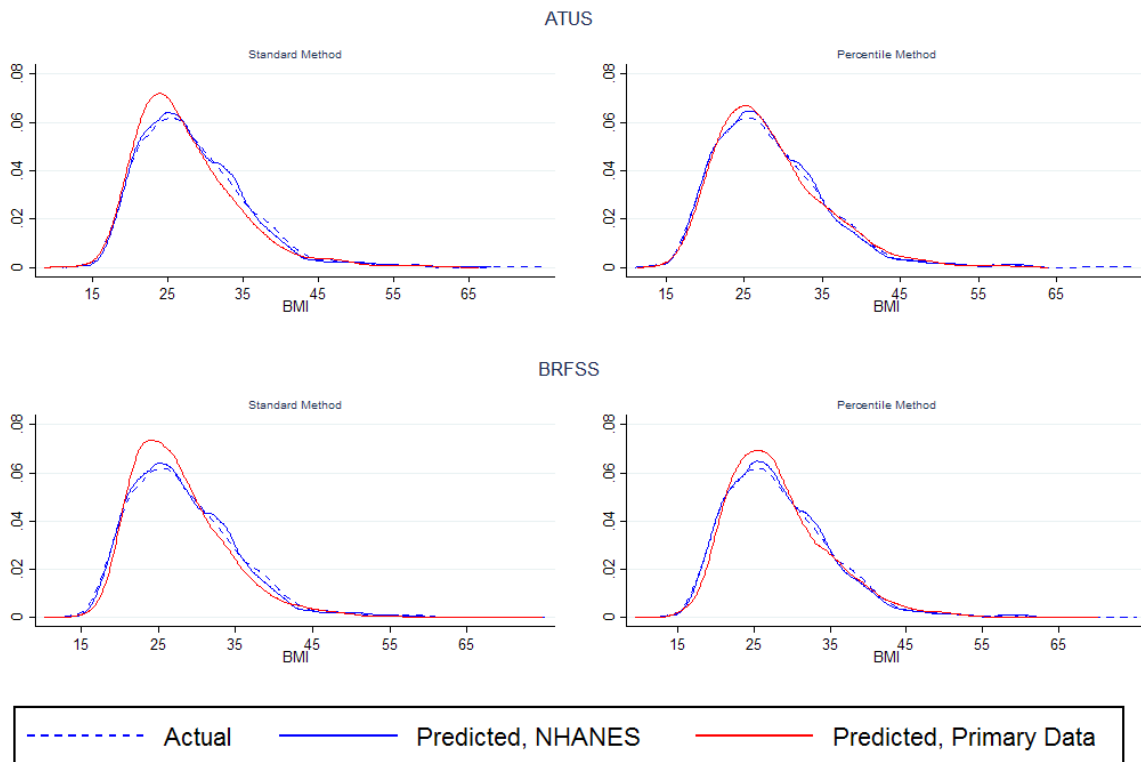|	Standard Method	Percentile Method
ATUS vs. NHANES	0.809	0.480
BRFSS vs. NHANES	0.819	0.382

Figure 2C. Standard vs. Percentile Prediction Methods for Black Women

Kolmogorov-Smirnov Tests (*p*-values) for Equality of Density Functions

	Standard Method	Percentile Method
ATUS vs. NHANES	0.049	0.734
BRFSS vs. NHANES	0.043	0.959

Figure 2D. Standard vs. Percentile Prediction Methods for Black Men

Kolmogorov-Smirnov Tests (*p*-values) for Equality of Density Functions

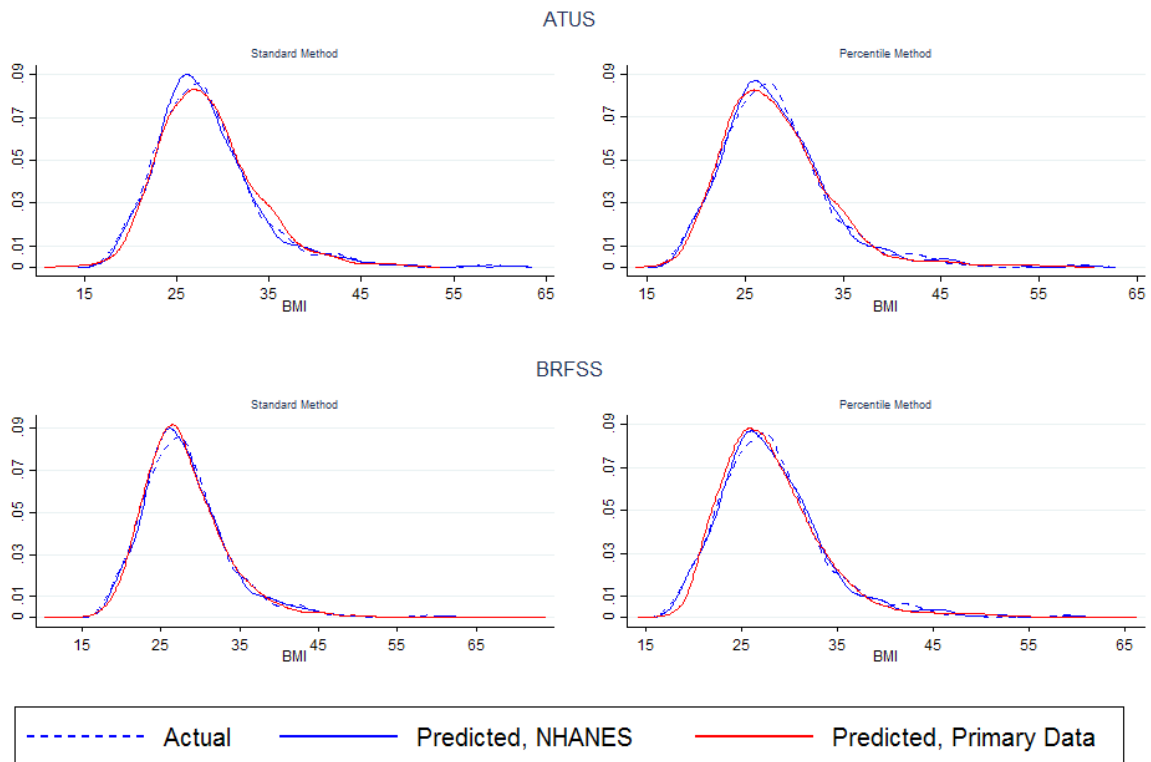|	Standard Method	Percentile Method
ATUS vs. NHANES	0.075	0.312
BRFSS vs. NHANES	0.001	0.158

Figure 2E. Standard vs. Percentile Prediction Methods for Other Women

Kolmogorov-Smirnov Tests (*p*-values) for Equality of Density Functions

	Standard Method	Percentile Method
ATUS vs. NHANES	0.001	0.263
BRFSS vs. NHANES	0.005	0.254

Figure 2F. Standard vs. Percentile Prediction Methods for Other Men

Kolmogorov-Smirnov Tests (*p*-values) for Equality of Density Functions

	Standard Method	Percentile Method
ATUS vs. NHANES	0.018	0.461
BRFSS vs. NHANES	0.813	0.579

Figure 3. Prevalence of Obesity by State Among Adults Ages 19-64
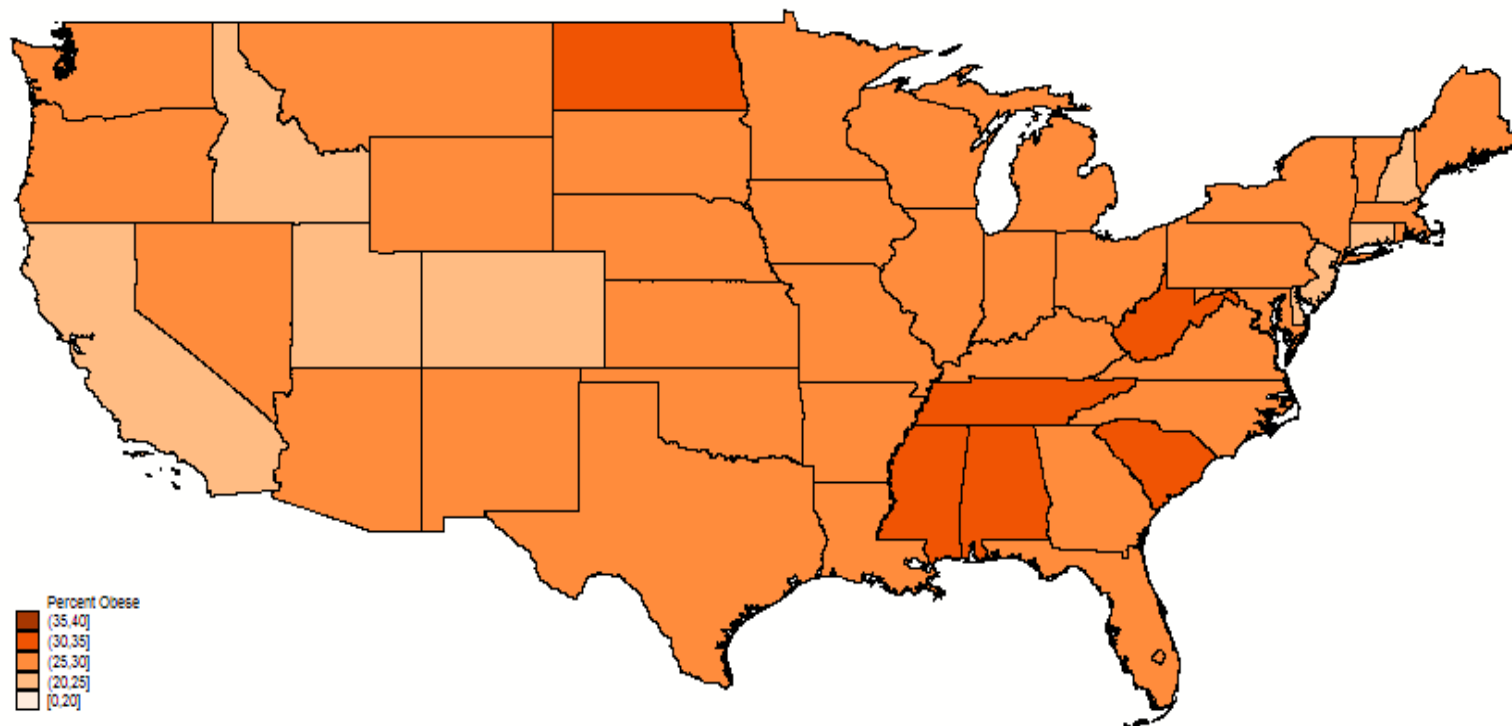Self-Reported BMI

Percent Obese
(35,40]
(30,35]
(25,30]
(20,25]
[0,20]

Figure 4. Prevalence of Obesity by State Among Adults Ages 19-64
Standard Validation Method

Percent Obese
(35,40]
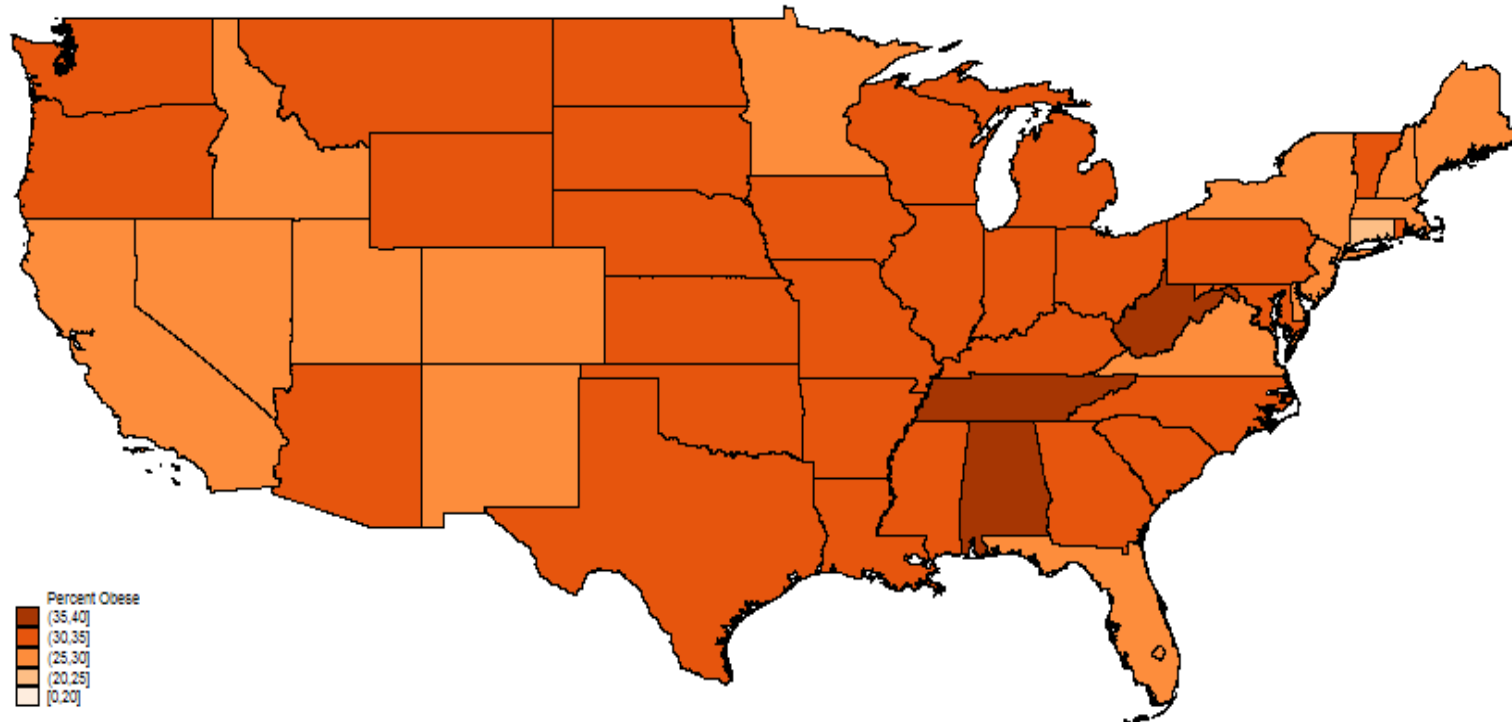(30,35]
(25,30]
(20,25]
[0,20]

Figure 5. Prevalence of Obesity by State Among Adults Ages 19-64
Percentile Validation Method