Reporting error in weight and height among the elderly: Implications and recommendations for estimating healthcare costs

by

Johanna Catherine Maclean
Department of Economics
Temple University
catherine.maclean@temple.edu

Asia Sikora Kessler
Department of Health Promotion
Social and Economic Behavioral Health
University of Nebraska Medical Center
asia.sikora@unmc.edu

Department of Economics
DETU Working Paper 15-01
January 2015

1301 Cecil B. Moore Avenue, Philadelphia, PA 19122

http://www.cla.temple.edu/economics/faculty/detu-working-paper-series/

Reporting error in weight and height among the elderly: Implications and recommendations for estimating healthcare costs

Johanna Catherine Maclean[1]*

Asia Sikora Kessler[2]

Abstract: A large literature has examined the healthcare consequences of obesity. A major barrier to careful study of these consequences is reliance on self-reported measures of weight and height. Previous research has developed algorithms to adjust for such error among working age adults. In this study we consider elderly adults, a group likely to differ in reporting error patterns from working age adults due to involuntary weight loss and changes in cognition, muscle mass, and bone density. We first provide evidence on the degree and type of reporting error in this population. Second, we consider how well standard approaches to adjusting for such error preform in an elderly population in terms of estimating obesity prevalence and regression coefficients. These findings have direct implications for evaluating anti-obesity programs among the elderly and estimating the obesity-related healthcosts to the Medicare program.

Key words: Healthcare costs; Medicare; reporting error; validation data; weight; height; obesity; elderly adults.

JEL classification: I1.

*Corresponding author.

Acknowledgements: We thank Lauren Nichols, Josh Pinkston, Michael Richards, Douglas Webber, session participants at the 2013 Southern Economic Association Annual Meetings and 2014 American Health Economists Society Conference for suggestions. All errors are our own.

[1] Department of Economics, Temple University, Philadelphia, PA. Email: catherine.maclean@temple.edu, Phone: 215-204-0560.
[2] Department of Health Promotion, Social and Economic Behavioral Health, University of Nebraska Medical Center, Omaha NE. Email: asia.sikora@unmc.edu, Phone: 402-552-7258.

1. Introduction

Obesity, or excess body fat, has generated a great deal of policy and public health concern due to its rapid rise and high prevalence, and the belief that it is linked with a range of costly economic and social consequences. As of 2012, 34.9% of adults in the United States are obese (Ogden, Carroll et al. 2014). Prevalence rates among adults 60 years and older are somewhat higher: 35.4%. Further, there is empirical evidence to support the belief that obesity leads to costly social and economic consequences. Obesity is the second leading cause of preventable death and contributes to a host of morbidities including Type II diabetes, asthma, cancer, and heart disease (Dixon 2010; Flegal, Kit et al. 2013) and estimates based on working age adults suggest that obesity raises healthcare costs by $2,741 annually per obese adult or $190.2 billion per year overall (Cawley and Meyerhoefer 2012).

Understanding the healthcare costs among the elderly important from a public finance perspective as the elderly have direct access to public health insurance through Medicare. The cost of Medicare to the U.S. was $585.7 billion in 2013, or 20% of total healthcare expenditures (Centers for Medicare & Medicaid Services 2014). Germane to our study, obesity is considered a major driver of Medicare costs (Finkelstein, Fiebelkorn et al. 2003; Thorpe and Howard 2006; Finkelstein, Trogdon et al. 2009). In particular, Finkelstein, Trogdon et al. (2009) document that in 2006 obese beneficiaries cost the Medicare program $600 more per year, on average, than normal weight beneficiaries. These costs were largely attributable to the use of prescription medications covered by Medicare Part D. Moreover, there is concern among health policy makers regarding the financial sustainability of the Medicare program (Baicker, Shepard et al. 2013; Davis, Schoen et al. 2013; Eibner, Goldman et al. 2013). Such concern is driven by increases in healthcare costs generally (Hartman, Martin et al. 2014) and, in particular, increases in healthcare expenditures in later stages of life (Alemayehu and Warner 2004). The increasing share of the population who are elderly, i.e. those individuals who utilize but do not contemporaneously pay into Medicare, adds to this concern. In 2013, 14% of the population was age 65 or older (authors' calculation based on the 2013 American Community Survey) and there were 51.3 million enrollees in the Medicare program (Hartman, Martin et al. 2014). Due to its high costs, developing and implementing effective programs to reduce obesity among the elderly, and in turn costs to the Medicare program, may be warranted.

A necessary condition for careful study of the healthcare costs attributable to obesity and evaluation of anti-obesity programs in the elderly population is the ability to accurately measure obesity in data. Ideally researchers would utilize physical measurements of both weight and height collected by trained medical professionals. These physical measurements would be combined to construct indicators for obesity (e.g., implied body mass index or BMI) and utilized in policy evaluations and econometric analyses. However, the majority of social science data sets (e.g., Medical Expenditure Panel Survey [MEPS], National Health Interview Survey [NHIS]) utilized to study healthcare utilization and costs collect respondent reports of weight and height. Self-reports are known to contain substantial, and importantly systematic or non-classical, reporting error (Rowland 1990; Cawley 2004; Cawley and Burkhauser 2006; Cawley, Maclean et al. 2014; Courtemanche, Pinkston et al. 2014). Reports are collected by social science surveys, given their well-known limitations, as they are less costly (both in terms of financial and hassle costs) than more accurate measurements of weight and height. To minimize reporting error in self-reports of weight and height, researchers have relied on statistical errors-in-variables theory (Lee and Sepanski 1995) and developed algorithms based on external validation data sets (e.g., National Health and Nutrition Examination Survey [NHANES]) that contain information on both reports and measurements (Cawley 2004; Cawley and Burkhauser 2006; Courtemanche, Pinkston et al. 2014). Applying these algorithms to reported weight and height has been shown to reduce reporting error. Although this work represents an important contribution to the literature on the economic and social consequences of obesity, it has focused on the working age population (e.g., ages 18 to 64 years). Thus the extent to which reporting error is present in elderly adults, and the ability of algorithms to minimize reporting error the elderly, is unclear. For example, elderly adults may be more likely than younger adults to misreport weight or height, and thus BMI or other measures of obesity, due to failing cognition, height shrinkage, involuntary weight loss, declining muscle mass and bone density, and other physical changes (Kuczmarski, Kuczmarski et al. 2001; Sahyoun, Maynard et al. 2008; Stommel and Schoenborn 2009). However, the elderly are also more likely to engage with healthcare professionals (Bernstein, Hing et al. 2003; Alemayehu and Warner 2004) and thus may be more likely to have recently had their weight and height professionally measured. While on net it seems that the elderly would be more likely to misreport weight and height, this prediction is not entirely clear *ex ante*.

The overarching goal of our study is to provide evidence on reporting error among elderly adults. Our study is not the first to document and examine the implications of reporting error in weight and height. Instead, we apply insight developed in previous studies based on samples of working age adults to a novel and policy relevant population. Specifically, our study has four objectives: 1) to provide new information on the extent of and characteristics of reporting error in elderly adult (65 years and older) self-reports of weight, height, and (implied) BMI; 2) apply previously developed methods to minimize reporting error in an elderly population; 3) tailor these previously developed methods to an elderly adult population to further minimize error; and 4) assess the impact of reporting error in weight and height on regression coefficients estimated in an elderly adult population. This information may be useful to researchers and policy makers interested in estimating healthcare costs attributable to Medicare and other payers, and in evaluation of anti-obesity policies that may affect the elderly (e.g., taxation of unhealthy foods).

2. Background on reporting error in weight and height

Bound, Brown et al. (2001) provide a comprehensive review of the consequences of error in economic variables. In their review, the authors state that economists tend to assume that reporting error is classical (i.e., that it is uncorrelated with the true value of the variable) and that its presence in a regressor implies that the associated regression coefficient estimate is biased toward zero. Under this assumption regression coefficients have the interpretation of lower bounds on the true relationship. However, Bound, Brown et al. (2001) show that this commonly held belief is only true under certain conditions. Specifically if the regression model is linear, there is a single regressor in the model, and the reporting error in the single regressor is classical (i.e., the error in the regressor is mean zero and independent of that regressor).

Many applications faced by economists interested in estimating the healthcare costs attributable to obesity likely do not fall into this category. In particular, while reporting error in a continuous variable such as weight, height, or implied BMI could be classical, error in binary variables such as clinical weight classifications (e.g., underweight, obesity) cannot be classical because error is negatively correlated with the true value (Cawley, Maclean et al. 2014). Bound et al document that the bias resulting from reporting error depends on a number of factors, such as the nature of the error, the type of regression model (linear or nonlinear), and whether the reporting error is correlated with other regressors. Specifically, the authors argue that the direction of bias introduced by reporting error is difficult to sign *ex ante* in many practical applications and reliance on standard heuristics (e.g., reporting error leads to attenuation bias) is likely to generate incorrect conclusions and/or interpretations of regression coefficients.

Importantly if economists attempting to estimate the costs attributable to obesity for payers (e.g., Medicare, private insurers) interpret their regression coefficients as lower bounds on the true obesity effects this interpretation may lead to the incorrect conclusion that obesity-related costs are *greater* than predicted by the model. This incorrect interpretation can, in turn, lead to an overinvestment in anti-obesity policies and programs. Such policies and programs may in fact be harmful, or at least ineffective, as excess weight may be protective for elderly adults (Miller and Wolfe 2008; Oreopoulos, Padwal et al. 2008; Kuk and Ardern 2009; Donini, Savina et al. 2012; Flegal, Kit et al. 2013); however, the literature remains somewhat mixed on this subject (Losonczy, Harris et al. 1995; Masters 2013).

Self-reported weight and height variables are known to contain systematic reporting error (Rowland 1990; Courtemanche, Pinkston et al. 2014). For example, Cawley (2002) shows that underweight individuals tend to over-report weight while overweight and obese individuals tend to under-report weight. Rowland (1990) documents that overweight persons are more likely to overstate their weight than individuals in lower levels of the weight distribution. Recent work by Cawley, Maclean et al. (2014) reports a range of statistical tests that suggest error in reported weight, height, and implied BMI is systematic and correlated with standard regressors. Pinkston (2014) shows that the interview method can influence a respondent's weight and height reporting. Members of the NLSY79 cohort who are interviewed over the telephone report lower weight, on average, than members who are interviewed face-to-face. Collectively, these studies suggest that reporting error in weight, height, and implied BMI is not classical and the assumption that regression coefficients based on reports are attenuated is not correct in many applications.

Self-reports are routinely collected in health and social science surveys as they are less costly (in terms of direct financial costs and indirect hassle costs) to collect than more accurate measures (Burkhauser and Cawley 2008). Researchers have developed algorithms to correct for reporting error (Cawley 2002; Cawley 2004; Cawley and Burkhauser 2006; Courtemanche, Pinkston et al. 2014) utilizing insight from statistical work on errors-in-variables and validation data (Lee and Sepanski 1995; Bound, Brown et al. 2001). For example, Cawley and Burkhauser

(2006), henceforth 'CB', develop an algorithm based on comparisons of reported weight and height, and measured weight and height in the NHANES III Survey (1988 and 1994). Specifically, CB estimate race- and sex-specific regressions of measured weight (height) on a quadratic in reported weight (height) and a quadratic in age. They use the coefficient estimates from these regressions to predict weight and height in the data sets that contain only self-reports. CB show that use of this algorithm can substantially reduce reporting error in body weight variables based on reports and improve weight status classification accuracy. This method has been widely used in studies investigating the economic and social costs of obesity (Courtemanche, Pinkston et al. 2014)

Courtemanche, Pinkston et al. (2014) (henceforth 'CPS') build on the work of CB and use percentile ranks rather than levels of reported weight and height in their prediction models. Consistent with CB, CPS use the NHANES as their validation data set. CPS delineate the necessary conditions for the CB approach to hold:

1) There exists a surrogate for the measured weight or height, reports in both CB and CPS. Reports are a surrogate for measurements if the distribution of an outcome variable y given reports and measurements is the same of the distribution of y given measurements only.

2) The surrogate (reports) must satisfy transportability across the validation and primary data sets. Transportability is typically described as the distribution of measurements conditional on reports being the same in both the validation and primary data sets.

CPS argue that the CB approach requires the assumption that the expectations of measured weight and height conditional on the reported values are the same in both the external validation data set (that contains both measurements and self-reports) and the primary data set (that contains self-reports only). Further, CPS state that their approach is robust to differences between the primary and validation samples in the severity (or type) of reporting error provided that the rankings of respondents based on reported values resembles the rankings based on actual measures in both data sets. For this reason, CPS argue that the transportability assumption may not hold in the CB approach. CPS show that their approach often outperforms the CB approach in several U.S. large-scale data sets in terms of accurately estimating regression coefficients. CPS acknowledge that, like CB, their approach does require that researchers assume that reports must serve as a valid surrogate for measurements.

These previous studies of reporting error in weight and height focus exclusively on working age adults (e.g., 18 to 64 years), however. In their comprehensive review, Bound, Brown et al. (2001) note the importance that the primary and validation data sets are drawn from the same population. We suspect that using a validation data set based on a working age population is problematic for a primary data set of elderly adults. As noted earlier in the manuscript the elderly may display different patterns of reporting error in weigh, height, and implied BMI than working age adults. In this study we aim to provide information on the degree, and type, of reporting error in weight, height, and implied BMI among the elderly.

3. Data, variables, and methods

3.1 Primary data set: Health and Retirement Study

Our primary data set is the Health and Retirement Study (HRS), a rich survey sponsored by the National Center for Aging and the Social Security Administration. The HRS surveys a national sample of more than 32,000 non-institutionalized Americans over the age of 50, biennially since 1992. The survey is designed to assess changes in health and labor force participation in this population (Juster and Suzman 1995).

We focus specifically on elderly adults (65+) who completed the HRS Enhanced Face-to-Face interview (Crimmins, Guyer et al. 2008). In the 2006 to 2012 rounds a random sample of one half of households were pre-selected for the Enhanced Face-to-Face interview which included physical measurements of weight and height in addition to several other measurements (i.e., blood pressure, lung function, handgrip strength, balance tests, timed walk, and weight circumference). HRS interviewers were provided training measurement procedures for these outcomes to ensure accurate measurements. See Crimmins, Guyer et al. (2008) for specific details on weight and height collection procedures. Respondent selections were made at the household level to ensure that the same request was made to all household members participating in the survey. The enhanced interview was conducted at the same time as the main HRS interview, thus minimizing time differential between weight and height measurements and reports. Moreover, HRS respondents were not aware that they were preselected for the enhanced interview prior to the interview (authors' personal communications with HRS administrators). These attributes of the enhanced interview minimize concerns regarding behavioral response by respondents, e.g., weight loss prior to the interview (Han, Norton et al. 2009).

We use this measured weight and height information to construct our body weight measures (weight, height, and BMI, underweight, and obesity implied by weight and height). Administrators included information in the survey on the setting under which measurements was taken (e.g., type of flooring). We exclude those respondents who wore shoes during weight or height measurement, and/or were measured (weight or height) on high carpet. We merge in comparable information on reported body weight variables (weight, height, and implied BMI, underweight, and obesity implied by weight and height), in combination with information on healthcare utilization and demographics from the HRS core interviews contained in the RAND Fat Files Version L. After making several exclusions, described in a later section of the manuscript, our final HRS primary analysis data set includes 13,844 elderly individuals ages 65 years and older (5,793 men and 8,051 women).

3.2 External validation data set: National Health and Nutrition Examination Survey (NHANES)

We follow the majority of studies that develop algorithms to minimize error in self-reported weight, height, and implied BMI, and utilize the NHANES as our external validation data set. The NHANES is sponsored by the National Center for Health Statistics of the Centers for Disease Control and Prevention and surveys a nationally representative sample of the U.S. civilian non-institutionalized population. Specifically, we use the Continuous NHANES 2005/06, 2007/08, 2009/10, and 2011/12 to overlap as closely as possible with our HRS data. The public use Continuous NHANES are only available in two year intervals, thus we cannot more accurately match these data to our HRS analysis sample. The 2005 to 2012 Continuous NHANES data are comparable to the NHANES III utilized by CB, however they better match our HRS study period. As noted by Bound, Brown et al. (2001) transportability requires that the primary and validation data sets be representative of the same population. Matching our primary and external validation data sets in terms of time makes it more likely that transportability is achieved. Moreover, we utilize population weights in both the HRS and NHANES which are designed to generate nationally representative estimates to meet this assumption in our analysis.

In the NHANES respondents report their weight and height during a household interview, and are measured and weighed by medical professionals during a subsequent examination in a mobile examination center. A limitation of the Continuous NHANES is that, unlike the HRS enhanced interview, the measurement of weight and height does not immediately follow the individual's report of their weight and height. Instead, the measurement takes place an average of two to three weeks after the interview. Thus, a share of any difference between weight/height reports and measurements may not be error but instead may capture change in weight/height that occurred in the period between interviews (e.g., involuntary weight loss due to sickness, changes in diet or physical activity). Moreover, respondents know at the time reports are collected that they will be measured in the near future suggesting that behavioral responses (e.g., more truthful reporting of weight/height). We exclude respondents with missing weight and height (measured or reported), race, ethnicity, or sex. Our final NHANES external validation sample consists of 17,885 individuals ages 18 to 64 (8,680 men and 9,205 women) and 5,069 elderly individuals ages 65 years and older (2,529 men and 2,540 women).

3.2 Weight and height variables

We examine measures of weight (pounds) and height (inches), both physical measurements and reports. Following previous work (Cawley and Burkhauser 2006) we exclude respondents with measured height less than 48 inches and greater than 80 inches in our primary and external validation data sets. In addition to weight and height, we examine three measures of implied body weight: a continuous measure for BMI (i.e., weight in pounds*703/height in inches squared), and dichotomous indicators for underweight (BMI < 18.5) and obesity (BMI ≥ 30). Specifically, medical research links involuntary weight loss among older adults with increased risk of morbidity and mortality (Felson, Zhang et al. 1992; Newman, Yanez et al. 2001; Flegal, Graubard et al. 2007; Janssen 2007). We do not examine overweight ($25 \geq$ BMI > 30), as evidence points to moderate levels of excess weight being protective against mortality and morbidity, particularly in an older population, and the current definition of overweight based on BMI may be too restrictive for older adults (Heiat, Vaccarino et al. 2001).

We next describe the various approaches we utilize to address reporting error in our weight, height, and BMI variables. These approaches are based on work by Cawley and Burkhauser (2006) and Courtemanche, Pinkston et al. (2014), and were described in more detail in Section 2. We rely on external validation data (Continuous NHANES) as would be the case in most economic studies that relay social science data sets that do not contain measurements (e.g., MEPS, NHIS). We could have developed our correction algorithms using comparison between self-reports and measurements contained in the HRS. However, we chose to utilize the

NHANES to mimic the setting in which researchers do not have access to measurements in their primary data set and must instead rely on an external validation data set. It is not clear to us why researchers would use reports when measurements are available in the primary data set.

Recall that previous work by CB and CPS has developed these algorithms for the working age population (18-64 years), but they have not been tested in an elderly sample. To assess the performance of the CB algorithm in an elderly adult sample, we follow CB and regress measured height (weight) on a quadratic in reported height (weight) and a quadratic in age (measured in years) using a sample of adults ages 18 to 64 years in the 2005 to 2012 Continuous NHANES. We apply the corrections to the reported weight and height variables in the HRS. We refer to these variables as the 'CB working age corrections'. We next update the working adult correction algorithm developed by CB using a sample of adults age 65 years and older in the 2005 to 2012 Continuous NHANES. Transportability requires that the primary and external validation data sets are drawn from the same underlying population, thus using age-matched samples may allow us to meet this condition (Bound, Brown et al. 2001). We apply the elderly adult corrections to the reported weight and height variables in our HRS sample following a comparable procedure. We refer to these variables as the 'CB elderly corrections'.

We next develop measures based on the CPS approach to addressing reporting error in weight and height. In particular, this approach uses the percentile rank of a report to generate a predicted value. We mimic our earlier CB approach and preform this correction for 1) working age adults and 2) elderly adults in the 2005 to 2012 Continuous NHANES. We refer to these corrections as the 'CPS working age corrections' and 'CPS elderly corrections' respectively.

To summarize, we consider six alternative sets of weight and height variables: measurements, uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections. We treat the measurements as the best estimate of true weight and height in our analysis. All programs and parameters utilized in generating our corrected measures are available on request.

3.4 Healthcare utilization: Hospitalizations

We construct an indicator variable coded one if the respondent reported spending a night in the hospital in the two years prior to the HRS survey and zero otherwise. We focus on hospital admissions as they are a costly form of healthcare that can, in some cases, be avoided by appropriate use of preventive or primary care. A limitation of this variable is that it likely predates our weight, height, and BMI variables (which are measured/collected at the time of the HRS survey) for many individuals, in the extreme by two years. However, we assume for the purposes of our study that the weight, height, and BMI variables at the time of the survey are a reasonable proxy for those at the time of hospitalization. This scenario could be a particular problem for our analysis if the respondent's weight changed due to the health event that led to the hospitalization, however. We note this as a limitation of our study.

3.5 Other control variables used in regression models

In all regression models, we adjust for age (55 to 59 years, 60 to 64 years, 65 to 69 years, 70 to 74 years, 75 to 79 years, 80 to 84 years, and 85 and older, with 50 to 54 years as the omitted category), race/ethnicity (African American, Hispanic, and other race with white as the omitted category), educational attainment (high school, some college, and a college degree with less than high school as the omitted category), marital status (divorced, separated, widowed, never married, and cohabiting with married as the omitted category), and survey fixed effects (2008, 2010, and 2012 with 2006 as the omitted category). We exclude HRS respondents with missing demographics from our analysis sample.

3.6 Methods

We first perform a range of statistical tests to document the extent of reporting error in weight, height, and BMI in our elderly HRS sample. These analyses are conducted for two reasons: 1) to document the degree and type of reporting error, and 2) to assess whether reporting error (if present) departs from the assumptions of classical reporting error, a necessary condition for reporting error to attenuate regression coefficients.

First, we calculate the level and percent of reporting error in reports of weight, height, and BMI. To calculate the degree of reporting error in weight, height, and BMI we use the following formula: $Report_{ij} - Measure_{ij}$. Where *j*=weight, height, or implied BMI for individual *i*. Second, we test whether the reporting error is mean zero using two-sided *t*-tests. We conduct these tests separately for the full sample and for specific demographic groups (sex, race, and ethnicity) to assess heterogeneity. Third, in separate models we regress reporting error in weight, height, and BMI on standard demographics (age, sex, race/ethnicity, educational attainment, marital status, and survey fixed effects) using least squares. If reporting error is correlated with these variables, this violates a necessary condition for reporting-error to bias coefficients towards zero (see Section 2). We next apply unconditional quantile regression (Firpo, Fortin et al. 2009) to study how reporting error varies across the distribution of weight, height, and BMI. If reporting error is particularly severe in specific points of the distribution of weight, height, or BMI, this finding would further suggest that reporting error in these variables is not classical.

We estimate mean values of BMI and the prevalence rates of underweight and obesity in our sample (separately by sex) across our six approaches to addressing error in reported weight and height variables (measurements, uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, CPS elderly corrections). This analysis will shed light on the implications of reporting error for estimates of mean BMI, and underweight and obesity prevalence rates among the elderly. We calculate the percent for whom the use of self-reports or a correction algorithm rather than measurements leads them to be true positives (classified as underweight/obese using reported and measured variables), false positives (classified as underweight/obese using reported variables and classified as not underweight/obese using measured variables), true negatives (classified as not underweight/obese using reported and measured variables), and false negatives (classified as not underweight/obese using reported variables and classified as underweight/obese using measured variables) for underweight and obesity, where true status is judged using measurements. Examining misclassification in this manner can shed light on which observations are likely to be misclassified (i.e., error type). For example, are individuals incorrectly classified as underweight when they are not or are they incorrectly classified as non-obese when they are in fact obese?

To estimate the associations between body weight and the probability of a hospitalization we utilize the following equation:

(1) $$P(H_i = 1) = \Phi(\beta_0 + \beta_1 BW_i + X_i\beta_2 + S_t\beta_3)$$

where H_i is an indicator for a hospitalization in the two years prior to the HRS interview, BW_i is a measure of body weight (BMI, underweight, or obesity), X_i is a vector of personal characteristics (age, race, ethnicity, education, marital status, and survey fixed effects), and the β*'s* are parameters to estimate. We estimate probit models and report average marginal effects rather than beta coefficients for ease of interpretation. We calculate heteroskedasticity robust standard errors. We estimate models separately for men and women given sex differences in body weight among the elderly (Ogden, Carroll et al. 2014).

4. Results

4.1 Evidence of reporting error in weight and height among older adults

Table 1 reports the level and percent of reporting error in weight, height, and BMI for the full sample and demographic groups (sex, race, ethnicity). In general, respondents in the HRS sample, and consistent with work on working age adults, tend to under-report weight and over-report height, leading to and under-estimate of BMI. For example, on average HRS respondents under-report their weight by 2.5 pounds, over-report their height by 1.2 inches, and thus under-report their BMI by 1.4 units. Under-reports of weight are greatest among women (2.9 pounds) and lowest among Hispanics (0.04 pounds). In terms of height, over-reporting is greatest among other races (2.1 inches) and lowest among women (1.1 inches). The extent of BMI under-reporting is greatest among other races (1.49 units) and lowest among Hispanics (1.1 units).

We next test, using two-tailed *t*-tests, whether reporting error in weight, height, and BMI is mean zero (a necessary condition for classical reporting error). Results from this testing are reported in Table 2. We report only the *p*-values from the *t*-tests for brevity, but more details on the analysis are available on request. These tests suggest that reporting error is not mean zero: we reject the null of mean zero reporting error for all variables in all groups ($p \leq 0.05$ or better).

We plot reported vs. measured weight, height, and BMI for the full sample in Figures 1, 2, and 3 respectively. If there is no systematic reporting error in these variables we would expect values to track the 45 degree line closely (we include the 45 degree line in each figure for ease of interpretation). Instead, we see substantial departures from the 45 degree line for all variables. In terms of weight, we observe a larger share of observations below the 45 degree line which suggests that respondents are more likely to under-report weight (Figure 1). The pattern is reversed for height (Figure 2): we see a disproportionate share of observations lie above the 45 degree line which implies that respondents are more likely to over-report their height. As observed in Table 1, the asymmetrical misreporting in weight and height leads to under-reported BMI: a larger share of the BMI values fall below the 45 degree line (i.e., under-report BMI).

To further explore reporting error among elderly respondents in the HRS, we estimate unconditional quantile regression or UQR (Firpo, Fortin et al. 2009). Such models allow consistent estimates of treatment effects at virtually any quantile of the unconditional distribution (quantiles are points taken at regular intervals from the cumulative distribution function of a random variable) and may uncover heterogeneity in relationships between reporting error and measured values of weight, height, and BMI that is masked by least squares which estimates effects at the mean. Specifically, we regress the reporting error on the measured value, we conduct this exercise separately for weight, height, and BMI. We estimate associations at every 5th quantile between the 5th through 95th quantile of the distribution of weight, height, and BMI. To examine statistical significance, we rely on parametric bootstrapped standard errors with 400 repetitions. Results from this analysis are reported in Figures 4 (weight), 5 (height), and 6 (BMI). 95% confidence intervals are reported with a shaded grey around the parameter estimates. For comparison, we report results generated in least squares regressions.

Least squares regressions suggest a negative association between measured weight and reporting error: a 1 pound increase in weight is associated with a 0.03 pound increase in the absolute value of reporting error in weight ($p \leq 0.01$). This regression masks a substantial degree of heterogeneity across the weight distribution. UQR documents that the lower weight individuals display less error in their weight than higher weight individuals. In particular, among those respondents at the 95th quantile of the unconditional weight distribution a 1 pound increase in weight is associated with a 0.14 pound increase in reporting error in weight. All coefficients estimated using UQR are statistically distinguishable from 0 at $p \leq 0.01$.

In terms of height we observe a reversal in this pattern (Figure 5). Least squares regressions show that a 1 inch increase in measured weight is associated with a -0.20 inch reduction in reporting error in height ($p \leq 0.01$). UQR analyses reveal that taller individuals are less likely to over-report height. Among those at the 95th quantile of the height distribution, a 1 inch increase in measured height is associated with a -0.49 inch reduction in reporting error in height. Examination of BMI suggests that focusing on mean effects estimated by least squares masks heterogeneity in reporting error and higher BMI individuals display more reporting error than lower BMI individuals (Figure 6). The finding that reporting error is not constant across the distribution of weight, height, and BMI provides further evidence that reporting error in these variables is not classical.

We next estimate least squares regressions of reporting error in weight, height, and BMI on demographics typically included in healthcare utilization regressions (i.e., age in years, race/ethnicity, educational attainment, marital status, and survey fixed effects). We estimate models separately for men and women. The analyses suggest that many of these standard regressors (i.e., age, African American race, Hispanic ethnicity, educational attainment, being widowed or divorced) are correlated with reporting error, and correlations vary across both sex and outcome (weight, height, BMI). Collectively, these results provide further evidence that reporting error in weight, height, and BMI is not classical.

4.2 Implications of reliance on reports for prevalence estimates

After providing several pieces of information that seem to conflict with necessary conditions for classical reporting error in weight, height, and BMI we next examine the bias in estimates of mean BMI and prevalence rates for two clinically important weight categories: underweight (BMI < 18.5) and obesity (BMI ≥ 30). In Table 4 we report estimates separately for men and women across measurements and our alternative approaches to addressing error in reports (uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections). We also calculate differences between estimates based on measurements and the alternative approaches to accounting for reporting error, using a non-parametric bootstrap (400 repetitions) to examine statistical significance.

Among men, relying on reports (whether uncorrected or corrected with any of the approaches we consider here) underestimates mean BMI. Based on measurements (which we treat as the true measurement) the mean BMI among men in our sample is 29.04. Reliance on uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections generates an estimate of mean BMI of 27.65, 28.44, 28.49, 28.62, and 28.59. These estimates understate mean BMI (relative to measurements) by 4.78%, 2.06%, 1.89%, 1.44%, and 1.63%. Among women the sample mean BMI based on measured weight and height is 28.71. All values based on reports (corrected or uncorrected) understate mean BMI by 0.90% to 5.08%. One exception is CPS working age corrections which *overstate* mean BMI among women by 1.93%. Summarizing across the alternative approaches to addressing reporting error, reliance on uncorrected reports leads to the largest understatement of BMI among both men and women. Moreover, in many cases the estimates generated using reported variables are statistically different from estimates generated using measured variables. The CPS approach appears to generate estimates closer to the true value than the CB approach, but reliance on an age-matched correction method (i.e., elderly CPS corrections) does not outperform a working age method.

Turning to estimates of underweight prevalence rates, reliance on reported weight and height tends to overstate underweight prevalence among men. 0.5% of the sample is classified as underweight using measurements while 0.70%, 0.60%, 0.60%, 1.1%, and 0.8% of men meet this classification based on uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections respectively. Thus, unlike the case of BMI, the CB approach (using either a working age population or an elderly population in the NHANES, our validation data set) out-preforms the CPS approach. Among women, 1.5% of the sample is underweight based on measured weight and height. Relying on reports leads to under- and over-estimates of the prevalence rate for underweight and reliance on uncorrected reports provides the most inaccurate estimate. The CB approach based on an age-matched sample in the validation data set offers the most accurate estimate of underweight prevalence among women. Differences between prevalence rates estimated using measurements and the alternative approaches to accounting for reporting error are often statistically different from zero, particularly among elderly women.

In the case of obesity, we find substantial bias in the estimates of the obesity prevalence rate among men if reports are utilized (either uncorrected or corrected). The obesity prevalence rate based on measurements among men is 37.7%. Reliance on uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections suggests that the prevalence rate is 26.5%, 32.1%, 32.3%, 33.4%, and 33.0%. Thus, reliance on these approaches to addressing reporting error (which includes no correction) understates true obesity prevalence among men by 29.7%, 14.9%, 14.3%, 11.4%, and 12.5% respectively (these differences are often statistically distinguishable from zero as well). Uncorrected reports produce the most biased estimate of obesity prevalence, while the CPS approach appears to perform better than the CB approach. Interestingly, using an age-matched sample in the validation data set does not appear to improve the estimate of the prevalence rate in the CPS approach although it does in the CB approach. Among women, the obesity prevalence rate based on measurements is 36.6%. As is the case among elderly men, use of reports (regardless of whether a correction approach is employed) leads to an under-estimate of obesity prevalence and the differences between estimates generated using measurements and reports are often statistically distinguishable from zero. Reliance on uncorrected reports yields the most biased estimate of this prevalence rate. However, use of either the CB or CPS approach substantially reduces the error in the estimate of the prevalence rate. As is the case among men, the CPS approach out-preforms the CB approach, although use of an age-matched sample in the validation data set does not lead to a reduction in bias.

Collectively these results present a quandary: although it is clear that correcting reports through either the CB or CPS approach reduces bias in estimates of mean BMI, and the prevalence rate of underweight and obesity in most cases, neither method strictly dominates the other. Moreover, using an age-matched sample in the validation data set does not unambiguously improve the quality of our estimates. These findings may be troubling for empirical researchers.

Next, we calculate true positives, false positives, true negatives, and false negatives for underweight (Table 5A) and obesity (Table 5B) status. That is, we compare underweight/obesity status using measurements and various approaches to addressing reporting error. This analysis provides some insight on the degree, and type, of misclassification when relying on reports. (We do not consider BMI here as these measures are not applicable to a continuous variable.)

In terms of underweight status, we find that less than 1% of men 2% of women are misclassified (in terms of either false positive or false negatives) when comparing reports (either uncorrected or corrected) to measurements. Among men, the CB approach (regardless of the age match between the HRS sample and the validation sample) generates a slightly higher share of correctly classified (either as underweight or not underweight) observations than uncorrected reports or the CPS approach. Among women, the results are more mixed. In particular, reliance on uncorrected reports generates a lower share of false negatives than correcting the reports.

Turning to obesity, among both men and women correcting reports using any of the approaches we consider increases the share of true positives. The improvement is substantial: roughly 5 percentage points among both men and women. Interestingly, reliance on uncorrected reports generates the *lowest* share of false positives among both men and women. Again, the improvement is non-trivial. For example, among men 1% of men are classified as obese using uncorrected reported variables and classified as not obese using the measured variables while 2.0%, 2.1%, 2.9%, and 2.3% of men fall into this group using CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections respectively. In terms of true negatives (classified as obese using both reports and measurements), uncorrected reports again outperform correction approaches among both men and women. The improvement in estimate quality ranges from 1 to 1.9 (1.5 to 4.3) percentage points among men (women). Lastly, we consider false negatives (those respondents who are classified as not obese occurring to reports and obese occurring to measurements). Among both men and women using some form of correction method substantially reduces the share of false negatives. For example, among men the share of false negatives is 12.2% using reports and 7.6%, 7.5%, 7.2%, and 7.0% using CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections. Among both groups, the CPS approach outperforms the CB approach in terms of reducing the share of false negatives. However, using an age-matched sample in the validation data set only improves classification in terms of false negatives among men.

4.3 Implications of reliance on reports for regression coefficients

In our final exercise, we compare regression coefficients using measurements, uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections in models of healthcare utilization and costs (as proxied by risk of a hospitalization). We bootstrap (with a non-parametric bootstrap using 400 repetitions) the difference between marginal effects estimated using measurements and the alternative approaches to accounting for reporting error to assess whether any differences are statistically significant. Results are reported in Table 6. In our analysis sample 31.3% of men and 29.1% of women report a hospitalization in the 2 years prior to the HRS interview. These regressions estimate associations, not causal effects. We defer estimation of the causal effects of body weight to further research and instead focus our attention on implications of reporting error.

In terms of BMI, we find that the coefficient estimates generated using reports (corrected or uncorrected) do not depart substantially from estimates generated using measurements among men. The coefficients range from 0.004 to 0.005, suggesting that a 1 unit increase in BMI is associated with a 1.3% to a 1.6% increase in the risk of a hospitalization. Among women, we find a reasonably tight range of regression coefficient estimates from 0.003 to 0.004, suggesting that a 1 unit increase in BMI is associated with a 1.0% to 1.4% increase in risk for a hospitalization. Differences between marginal effects estimated using measurements and reports are often statistically distinguishable from zero, particularly among women. The CB and CPS approaches appear to preform equally well in estimating regression coefficients, however, use of an age-matched sample in the validation data set does not appear to reduce bias.

Among men reliance on reports, regardless of whether a correction approach is adopted, appears to overstate the precision and/or magnitude of the association between underweight and risk of a hospitalization. The regression coefficient estimated using reports is 0.145 and is statistically indistinguishable from zero. However, 4 of 5 regression coefficients estimated using reports are statistically different from zero ($p \leq 0.10$ level or better) and the coefficient estimates range from 0.098 to 0.201. No approach to correcting error in reports preforms particularly well. Among women, all coefficients (measurements and reports) are statistically indistinguishable from zero. The estimate based on measurements is -0.008. We observe a wide range in the coefficient estimates based on reports, -0.008 to 0.058, and only one coefficient carries the correct (negative) sign. Differences between marginal effects estimated using measurements and reports are more likely to be statistically different among women than among men.

In terms of obesity, we find that reliance on reports (uncorrected or corrected) leads to an *overestimate* in the association between obesity status and risk of a hospitalization, and the differences between marginal effects estimated using measures and reports are often statistically distinguishable from zero. Among men, the estimate based on measurements is 0.043 while the estimates based on reports range from 0.042 to 0.054. Unexpectedly, reliance on uncorrected reports generates the *least* biased estimate (0.042). Use of an age-matched validation sample does not reduce bias in the estimated coefficients. Turning to women, we find even larger differences between the estimates generated using measurements and those generated using reports. The coefficient estimate is 0.036, 0.057, 0.046, 0.051, 0.044, and 0.045 using measurements, uncorrected reports, CB working age corrections, CB elderly corrections, CPS working age corrections, and CPS elderly corrections. All estimates using reports overstate the association between obesity and risk of a hospitalization. The difference ranges from 22.2% to 58.3%. Uncorrected reports produce the estimate with the greatest bias while the CB elderly correction produces the estimate with the least bias.

6. Discussion

The contributions of this study are first to measure the extent and consider the implications of reliance on reported weight, height, and implied variables in an elderly population, and second to examine how well approaches to addressing reporting error based on evidence from the working age population apply to elderly adults (Cawley 2002; Cawley 2004; Cawley and Burkhauser 2006; Courtemanche, Pinkston et al. 2014). We contend here that an elderly population, for myriad reasons, may warrant a separate set of correction algorithms.

Several findings emerge from our analysis. First, we document that, consistent with the working age adult population, there is substantial degree of reporting error in weight, height, and implied BMI among elderly adults. Second, statistical testing suggests that the reporting error in these variables is not likely classical and thus the assumption that reporting error in weight, height, and BMI among elderly adults should attenuate coefficient estimates in healthcare utilization regressions is not supported by the data. Third, reporting error in weight, height, and BMI leads to biased estimates of population level BMI, and prevalence rates for underweight and obesity. Fourth, in healthcare utilization regression models coefficients that rely on reports of weight and height often depart from regression coefficients generated using measurements of these variables. In particular, estimates based on reports very often overstate associations between body weight and risk of a hospitalization. Fifth, and perhaps most important, use of standard approaches to addressing reporting error do not necessarily eliminate, or even reduce, bias in estimates of means, prevalence rates, or regression coefficients. Moreover, no approach emerges as the dominant approach. Instead, the preferred approach appears to be specific to the measure of obesity (BMI, underweight, or obesity), type of error (population mean, false positive, regression coefficient), and sample (men vs. women). For this reason, we cannot recommend a particular approach to researchers interested in estimating the healthcare utilization or costs attributable to obesity among the elderly.

This conclusion is supported even when the approaches are tailored to an elderly sample using the NHANES, matched closely in terms of timing, as an external validation data set. That matching on age between the primary and external validation data sets is somewhat surprising given that Bound, Brown et al. (2001) note the importance of drawing these data sets from the same population to ensure that transportability is satisfied. In this study we focus on regressions that estimate associations between measures of body weight and a costly form of healthcare utilization. That is, we consider implications of reporting error when measures of body weight are right hand side variables. An important extension to our work is rigorously considering the implications of reporting error in policy evaluations of anti-obesity policies/interventions (e.g., taxation of unhealthy foods such as sugar sweetened beverages). Such studies examine body weight as dependent variables in regression models.

Although our study is not a policy evaluation *per se*, it does have implications that are directly related to policy discussions around obesity among the elderly and financing of the

Medicare program. Previous work suggests that obesity is a major driver of Medicare costs (Finkelstein, Fiebelkorn et al. 2003; Thorpe and Howard 2006; Finkelstein, Trogdon et al. 2009). Thus policies and programs that address obesity among the elderly, and in turn contain Medicare costs, are potentially warranted. However, to understand the potential benefits of such programs and policies, we must first be able to accurately measure obesity. Our study speaks directly to the empirical challenges researchers may face in pursuit of this information.

Based on our findings, we encourage researchers, whenever possible, to utilize measured weight and height when examining body weight-related questions among elderly adults. When measured information is not available, we suggest that researchers compare estimates based on alternative correction methods to assess sensitivity in their findings. Moreover, we also recommend that survey administrators attempt to, even for a sub-set of respondents as in the Health and Retirement Study Enhanced Face-to-Face Interview, collect physical measurements of respondents' weight and height to facilitate careful analysis of the healthcare, and other, consequences of obesity. These questions will likely increase in their importance as the population continues to age and costs of supporting the Medicare program increase.

Table 1. Raw reporting error in weight, height, and BMI among the elderly: Health and Retirement Study 2006-12

Group	N	Weight (pounds)	Height (inches)	BMI (units)
Full sample	13,844	-2.454	1.225	-1.430
Men	5,793	-1.886	1.367	-1.390
Women	8,051	-2.875	1.120	-1.459
White	11,620	-2.557	1.189	-1.430
African American	1,748	-1.602	1.260	-1.400
Other race	476	-1.752	2.068	-1.493
Hispanic	991	-0.042	1.935	-1.142

Notes: HRS sample weights applied.

Table 2. Test for mean-zero reporting error in weight, height, and BMI among the elderly: Health and Retirement Study 2006-12

Raw error	Sample size	Weight	Height	BMI
Group	**N**	**Pounds**	**Inches**	**Units**
Full sample	13,853	0.0000	0.0000	0.0000
Men	5,795	0.0000	0.0000	0.0000
Women	8,058	0.0000	0.0000	0.0000
White	11,625	0.0000	0.0000	0.0000
African American	1,752	0.0000	0.0000	0.0000
Other race	476	0.0001	0.0000	0.0000
Hispanic	991	0.0103	0.0000	0.0000

Notes: *t*-tests applied, *p*-value from two-tailed tests reported.

Table 3A. Correlates of reporting error among elderly men: Health and Retirement Study 2006-12

Outcome:	**Weight (pounds)**	**Height (inches)**	**BMI (units)**
Sample mean	-1.886	1.367	-1.390
Age (years)	0.040**	0.034***	-0.016***
	(0.020)	(0.005)	(0.005)
African American	1.411***	0.077	0.123
	(0.493)	(0.110)	(0.120)
Other race	-0.382	0.671	-0.083
	(0.944)	(0.715)	(0.320)
Hispanic	3.293*	0.121	0.603*
	(1.859)	(0.330)	(0.351)
High school	-1.464***	-0.016	-0.290***
	(0.530)	(0.097)	(0.111)
Some college	-2.564***	-0.072	-0.427**
	(0.853)	(0.161)	(0.200)
College degree	-2.466***	-0.065	-0.329***
	(0.532)	(0.124)	(0.121)
Divorced/separated	0.857*	-0.028	0.175
	(0.503)	(0.104)	(0.115)
Widowed	-0.310	0.085	-0.162
	(0.382)	(0.101)	(0.108)
Never married	1.456	0.008	0.239
	(0.909)	(0.162)	(0.217)
N	5,793	5,793	5,793

Notes: All models control for survey fixed effects. Heteroskedasticity robust standard errors are reported in parentheses. Regressions estimated with OLS. HRS sample weights applied.
***; **; and *=statistically different from zero at the 1%; 5%; and 10% confidence level.

Table 3B. Correlates of reporting error among elderly women: Health and Retirement Study 2006-12

Outcome:	**Weight (pounds)**	**Height (inches)**	**BMI (units)**
Sample mean	-2.875	1.120	-1.459
Age (years)	0.122***	0.036***	0.002
	(0.015)	(0.008)	(0.006)
African American	0.489	0.038	-0.027
	(0.391)	(0.120)	(0.117)
Other race	0.256	0.582	-0.225
	(1.014)	(0.753)	(0.276)
Hispanic	1.622***	0.890**	0.117
	(0.499)	(0.437)	(0.192)
High school	-0.146	-0.252*	0.050
	(0.287)	(0.130)	(0.085)
Some college	-1.006**	-0.427***	0.011
	(0.460)	(0.165)	(0.152)
College degree	-0.109	-0.359**	0.193*
	(0.334)	(0.174)	(0.103)
Divorced/separated	0.366	0.020	-0.045
	(0.466)	(0.104)	(0.123)
Widowed	-0.332	0.038	-0.153**
	(0.231)	(0.093)	(0.072)
Never married	-0.186	0.952	-0.434
	(0.608)	(0.888)	(0.296)
N	8,051	8,051	8,051

Notes: All models control for survey fixed effects. Heteroskedasticity robust standard errors are reported in parentheses. Regressions estimated with OLS. HRS sample weights applied.
***; **; and *=statistically different from zero at the 1%; 5%; and 10% confidence level.

Table 4. Comparison of BMI, underweight, and obesity by alternative corrections for measurement error in self-reported weight and height variables among the elderly: Health and Retirement Study 2006-2012

	Men	**Women**
	Mean/proportion (*p*-value on difference between measures and reports*)	Mean/proportion (*p*-value on difference between measures and reports*)
Outcome: BMI		
Measurements	29.041	28.714
Uncorrected reports	27.651 (0.000)	27.255 (0.000)
CB working age corrections	28.444 (0.000)	28.200 (0.000)
CB elderly corrections	28.491 (0.000)	28.315 (0.000)
CPS working age corrections	28.622 (0.000)	29.269 (0.000)
CPS elderly corrections	28.567 (0.000)	28.455 (0.000)
Outcome: Underweight		
Measurements	0.005	0.015
Uncorrected reports	0.007 (0.152)	0.024 (0.000)
CB working age corrections	0.006 (0.507)	0.019 (0.008)
CB elderly corrections	0.006 (0.863)	0.016 (0.772)
CPS working age corrections	0.011 (0.001)	0.011 (0.003)
CPS elderly corrections	0.008 (0.070)	0.020 (0.004)
Outcome: Obese		
Measurements	0.377	0.366
Uncorrected reports	0.265 (0.000)	0.274 (0.000)
CB working age corrections	0.321 (0.000)	0.336 (0.000)
CB elderly corrections	0.323 (0.000)	0.341(0.000)
CPS working age corrections	0.334 (0.000)	0.380 (0.000)
CPS elderly corrections	0.330 (0.000)	0.345 (0.000)
N	5,793	8,051

Notes: HRS sample weights applied.
**p*-values on difference between measures and alternative approaches to accounting for reporting error in weight, height, and BMI. Non-parametric bootstraps (using 400 repetitions) applied.

Table 5A. Proportion of older adults classified as true positive, false positive, true negative, and false negative in terms of underweight (BMI<18.5) among the elderly: Health and Retirement Study 2006-12

Measurements vs.	**True positive**	**False positive**	**True negative**	**False negative**
Sample: Men				
Uncorrected reports	0.003	0.004	0.990	0.002
CB working age corrections	0.003	0.003	0.992	0.002
CB elderly corrections	0.003	0.003	0.992	0.003
CPS working age corrections	0.003	0.008	0.986	0.002
CPS elderly corrections	0.003	0.005	0.990	0.002
N	5,793			
Sample: Women				
Uncorrected reports	0.011	0.013	0.972	0.004
CB working age corrections	0.011	0.008	0.976	0.005
CB elderly corrections	0.010	0.006	0.979	0.006
CPS working age corrections	0.006	0.005	0.979	0.010
CPS elderly corrections	0.011	0.009	0.976	0.005
N	8,051			

Notes: True positives are classified as underweight using reported and measured variables. False positives are classified as underweight using reported variables and classified as not underweight using measured variables. True negatives are classified as not underweight using reported and measured variables. False negatives are observations classified as not underweight using reported variables and classified as underweight using measured variables. HRS sample weights applied.

Table 5B. Proportion of older adults classified as true positive, false positive, true negative, and false negative in terms of obese (BMI≥30) among the elderly: Health and Retirement Study 2006-12

Measurements vs.	**True positive**	**False positive**	**True negative**	**False negative**
Sample: Men				
Uncorrected reports	0.255	0.010	0.613	0.122
CB working age corrections	0.301	0.020	0.603	0.076
CB elderly corrections	0.302	0.021	0.602	0.075
CPS working age corrections	0.305	0.029	0.594	0.072
CPS elderly corrections	0.307	0.023	0.600	0.070
N	5,793			
Sample: Women				
Uncorrected reports	0.265	0.008	0.626	0.100
CB working age corrections	0.313	0.023	0.611	0.053
CB elderly corrections	0.315	0.026	0.608	0.051
CPS working age corrections	0.329	0.051	0.583	0.037
CPS elderly corrections	0.316	0.029	0.606	0.049
N	8,051			

Notes: True positives are classified as obese using reported and measured variables. False positives are classified as obese using reported variables and classified as not obese using measured variables. True negatives are classified as not obese using reported and measured variables. False negatives are observations classified as not obese using reported variables and classified as obese using measured variables. HRS sample weights applied.

Table 6. Associations between BMI, underweight, and obesity, and risk for a hospital admission among the elderly: Health and Retirement Study 2006-12

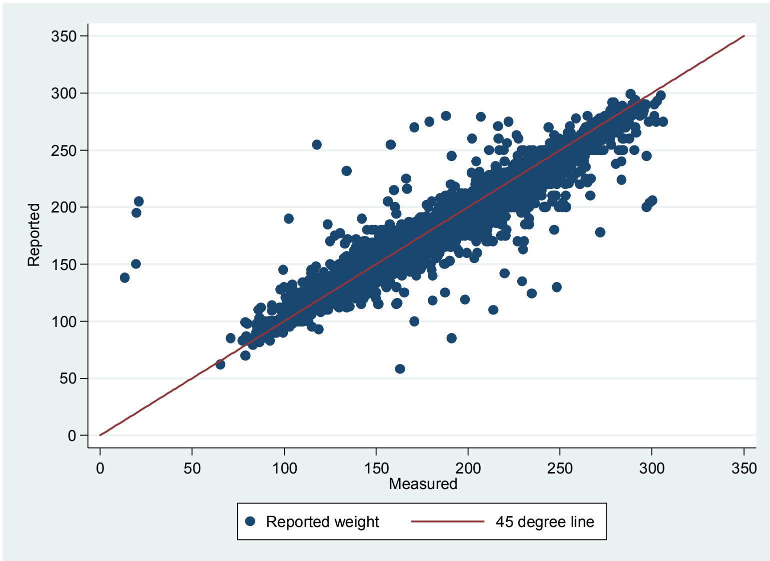
	Men	Women
Sample proportion reporting a hospital admission	0.313	0.291
Body weight measure: BMI		
Measurements	0.005***	0.003***
	(0.001)	(0.001)
Uncorrected reports	0.005***	0.004***
	(0.002)	(0.001)
	[0.778]	[0.007]
CB working age corrections	0.005***	0.004***
	(0.002)	(0.001)
	[0.698]	[0.101]
CB elderly corrections	0.005***	0.004***
	(0.002)	(0.001)
	[0.734]	[0.073]
CPS working age corrections	0.004***	0.004***
	(0.001)	(0.001)
	[0.149]	[0.000]
CPS elderly corrections	0.005***	0.004***
	(0.001)	(0.001)
	[0.636]	[0.096]
Body weight measure: Underweight		
Measurements	0.145	-0.008
	(0.108)	(0.044)
Uncorrected reports	0.172**	0.028
	(0.079)	(0.035)
	[0.780]	[0.303]
CB working age corrections	0.152*	0.044
	(0.091)	(0.039)
	[0.939]	[0.157]
CB elderly corrections	0.098	0.012
	(0.095)	(0.046)
	[0.663]	[0.583]
CPS working age corrections	0.141**	0.058
	(0.062)	(0.048)
	[0.966]	[0.194]
CPS elderly corrections	0.201***	-0.008
	(0.077)	(0.040)
	[0.588]	[0.991]
Body weight measure: Obesity		
Measurements	0.043***	0.036***
	(0.015)	(0.012)
Uncorrected reports	0.042***	0.057***
	(0.016)	(0.013)
	[0.966]	[0.012]
CB working age corrections	0.052***	0.046***
	(0.015)	(0.012)
	[0.297]	[0.138]
CB elderly corrections	0.050***	0.051***
	(0.015)	(0.012)
	[0.431]	[0.036]
CPS working age corrections	0.046***	0.044***
	(0.015)	(0.012)
	[0.776]	[0.269]
CPS elderly corrections	0.054***	0.045***

	(0.015)	(0.012)
	[0.253]	[0.227]
N	5,777	8,034

Notes: All regressions control for age, race/ethnicity, education, marital status, and survey fixed effects. Heteroskedasticity robust standard errors are reported in parentheses. *p*-values for difference between average marginal effects estimated in regressions using measurements and alternative approaches to accounting for reporting in weight, height, and BMI error reported in square brackets (non-parametric bootstraps using 400 repetitions applied). Regressions estimated with a probit model, average marginal effects are reported. HRS sample weights applied.

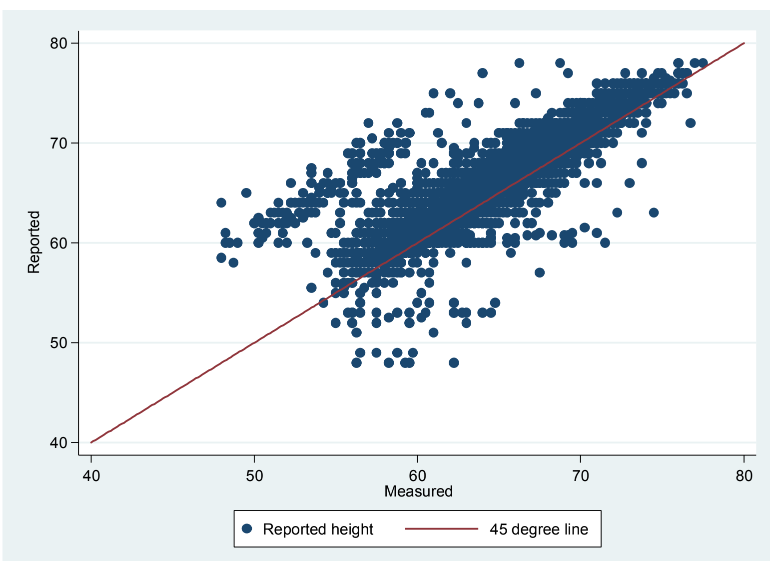
***; **; and *=statistically different from zero at the 1%; 5%; and 10% confidence level.

Figure 1: Reported vs. measured weight (pounds) among the elderly: Health and Retirement Study 2006-12



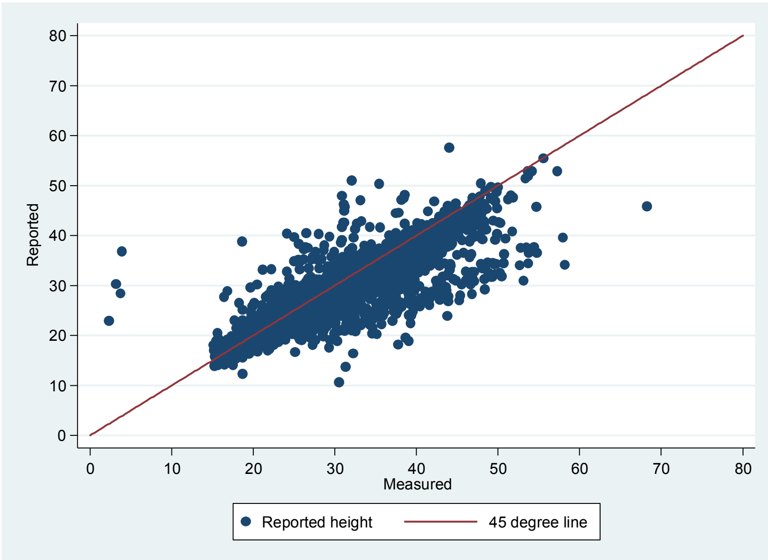
Notes: Full sample. Unweighted.

Figure 2: Reported vs. measured height (inches) among the elderly: Health and Retirement Study 2006-12



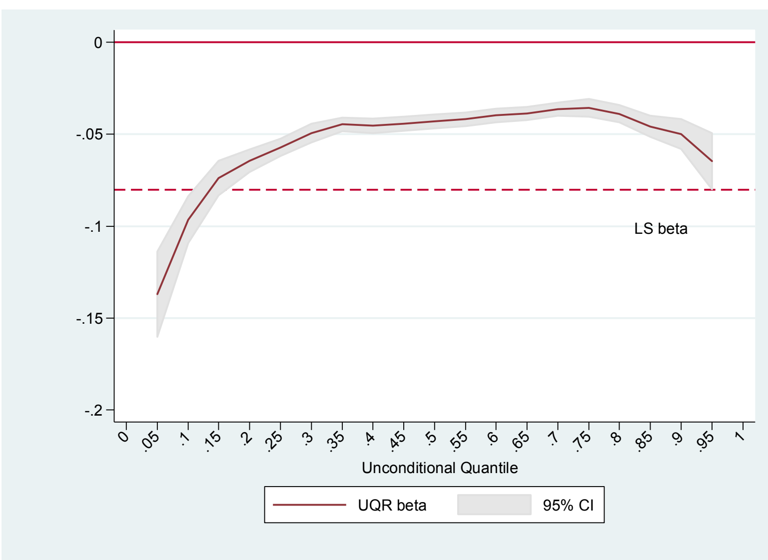
Notes: Full sample. Unweighted.

Figure 3: Reported vs. measured BMI (units) among the elderly: Health and Retirement Study 2006-12



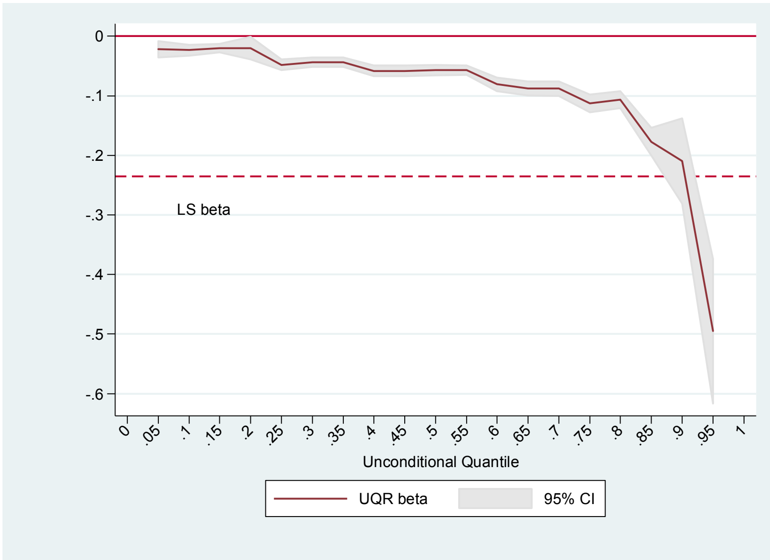
Notes: Full sample. Unweighted.

Figure 4: Heterogeneity in the association between measured weight and raw error in reported weight among the elderly: Health and Retirement Study 2006-12



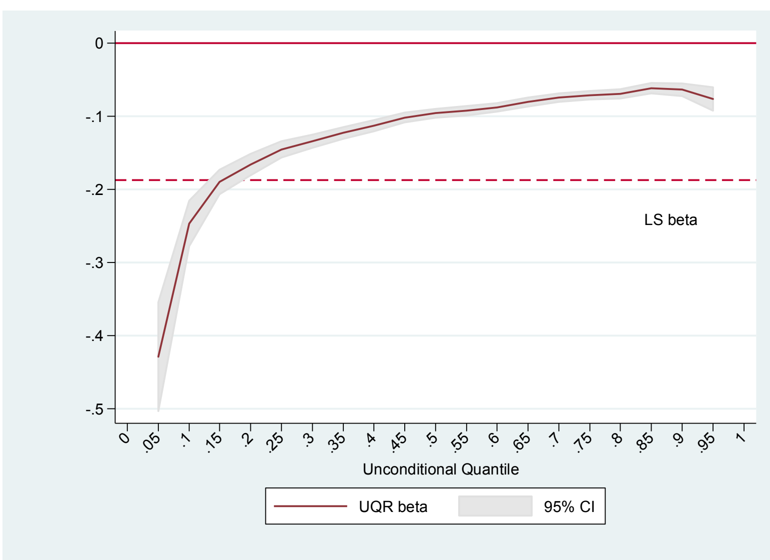
Notes: All regressions control for age, sex, race/ethnicity, education, marital status, and survey fixed effects. Models estimated with LS or unconditional quantile regressions. Heteroskedasticity robust standard errors are calculated in LS and non-parametric bootstrap (400 repetitions) standard errors are reported in unconditional quantile regressions. HRS sample weights applied.

Figure 5: Heterogeneity in the association between measured height and raw error in reported height among the elderly: Health and Retirement Study 2006-12



Notes: All regressions control for age, sex, race/ethnicity, education, marital status, and survey fixed effects. Models estimated with LS or unconditional quantile regressions. Heteroskedasticity robust standard errors are calculated in LS and non-parametric bootstrap (400 repetitions) standard errors are reported in unconditional quantile regressions. HRS sample weights applied.

Figure 6: Heterogeneity in the association between measured BMI and raw error in reported BMI among the elderly: Health and Retirement Study 2006-12



Notes: All regressions control for age, sex, race/ethnicity, education, marital status, and survey fixed effects. Models estimated with LS or unconditional quantile regressions. Heteroskedasticity robust standard errors are calculated in LS and non-parametric bootstrap (400 repetitions) standard errors are reported in unconditional quantile regressions. HRS sample weights applied.

References

Alemayehu, B. and K. E. Warner (2004). "The lifetime distribution of health care costs." Health Services Research **39**(3): 627-642.

Baicker, K., M. Shepard, et al. (2013). "Public Financing Of The Medicare Program Will Make Its Uniform Structure Increasingly Costly To Sustain." Health Aff (Millwood) **32**(5): 882-890.

Bernstein, A. B., E. Hing, et al. (2003). Health care in America: Trends in utilization. Hyattsville, MD, National Center for Health Statistics.

Bound, J., C. Brown, et al. (2001). Chapter 59 Measurement error in survey data. Handbook of Econometrics. J. J. Heckman and E. Leamer, Elsevier. **Volume 5:** 3705-3843.

Burkhauser, R. V. and J. Cawley (2008). "Beyond BMI: The value of more accurate measures of fatness and obesity in social science research." Journal of Health Economics **27**(2): 519-529.

Cawley, J. (2002). Addiction and the consumption of calories: Implications for obesity, Cornell University.

Cawley, J. (2004). "The Impact of Obesity on Wages." Journal of Human Resources **XXXIX**(2): 451-474.

Cawley, J. and R. V. Burkhauser (2006). Beyond BMI: The value of more accurate measures of fatness and obesity in social science research, National Bureau of Economic Research.

Cawley, J., J. C. Maclean, et al. (2014). Reporting Error in Weight and its Implications for Estimates of the Economic Consequences of Obesity.

Cawley, J. and C. Meyerhoefer (2012). "The medical care costs of obesity: an instrumental variables approach." Journal of Health Economics **31**(1): 219-230.

Centers for Medicare & Medicaid Services (2014). National Health Expenditure Accounts. Table 3 National Health Expenditures, by Source of Funds. Baltimore, MD.

Courtemanche, C., J. C. Pinkston, et al. (2014). Adjusting Body Mass for Measurement Error with Invalid Validation Data. NBER Working Paper Series. Cambridge, MA, National Bureau of Economic Research.

Crimmins, E., H. Guyer, et al. (2008). Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health and Retirement Study. Ann Arbor, MI, Survey Research Center University of Michigan.

Davis, K., C. Schoen, et al. (2013). "Medicare Essential: An Option To Promote Better Care And Curb Spending Growth." Health Aff (Millwood) **32**(5): 900-909.

Dixon, J. B. (2010). "The effect of obesity on health outcomes." Molecular and Cellular Endocrinology **316**(2): 104-108.

Donini, L. M., C. Savina, et al. (2012). "A Systematic Review of the Literature Concerning the Relationship between Obesity and Mortality in the Elderly." Journal of Nutrition Health & Aging **16**(1): 89-98.

Eibner, C., D. P. Goldman, et al. (2013). "Three Large-Scale Changes To The Medicare Program Could Curb Its Costs But Also Reduce Enrollment." Health Aff (Millwood) **32**(5): 891-899.

Felson, D. T., Y. Zhang, et al. (1992). "Weight loss reduces the risk for symptomatic knee osteoarthritis in women. The Framingham Study." Ann Intern Med **116**(7): 535-539.

Finkelstein, E. A., I. C. Fiebelkorn, et al. (2003). "National medical spending attributable to overweight and obesity: how much, and who's paying?" Health Aff (Millwood) **Suppl Web Exclusives**: W3-219-226.

Finkelstein, E. A., J. G. Trogdon, et al. (2009). "Annual medical spending attributable to obesity:

payer-and service-specific estimates." Health Aff (Millwood) **28**(5): w822-831.
Firpo, S., N. M. Fortin, et al. (2009). "Unconditional Quantile Regressions." Econometrica **77**(3): 953-973.
Flegal, K. M., B. I. Graubard, et al. (2007). "Cause-specific excess deaths associated with underweight, overweight, and obesity." Jama-Journal of the American Medical Association **298**(17): 2028-2037.
Flegal, K. M., B. K. Kit, et al. (2013). "Association of All-Cause Mortality With Overweight and Obesity Using Standard Body Mass Index Categories A Systematic Review and Meta-analysis." Jama-Journal of the American Medical Association **309**(1): 71-82.
Han, E., E. C. Norton, et al. (2009). "Weight and Wages: Fat Versus Lean Paychecks." Health Economics **18**(5): 535-548.
Hartman, M., A. B. Martin, et al. (2014). "National Health Spending In 2013: Growth Slows, Remains In Step With The Overall Economy." Health Aff (Millwood).
Heiat, A., V. Vaccarino, et al. (2001). "An evidence-based assessment of federal guidelines for overweight and obesity as they apply to elderly persons." Archives of Internal Medicine **161**(9): 1194-1203.
Janssen, I. (2007). "Morbidity and mortality risk associated with an overweight BMI in older men and women." Obesity **15**(7): 1827-1840.
Juster, F. T. and R. Suzman (1995). "An overview of the health and retirement study." Journal of Human Resources **30**: S7-S56.
Kuczmarski, M. F., R. J. Kuczmarski, et al. (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." J Am Diet Assoc **101**(1): 28-34; quiz 35-26.
Kuk, J. L. and C. I. Ardern (2009). "Influence of Age on the Association Between Various Measures of Obesity and All-Cause Mortality." Journal of the American Geriatrics Society **57**(11): 2077-2084.
Lee, L. F. and J. H. Sepanski (1995). "Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data." Journal of the American Statistical Association **90**(429): 130-140.
Losonczy, K. G., T. B. Harris, et al. (1995). "Does Weight-Loss from Middle-Age to Old-Age Explain the Inverse Weight Mortality Relation in Old-Age." American Journal of Epidemiology **141**(4): 312-321.
Masters, R. K. (2013). The Protective Effect of High Body Mass on Mortality Risk. Obesity Paradox or Data Artifact? Institute of Behavioral Science Population Program Working Paper Series. Denver, CO, University of Colorado.
Miller, S. L. and R. R. Wolfe (2008). "The danger of weight loss in the elderly." Journal of Nutrition Health & Aging **12**(7): 487-491.
Newman, A. B., D. Yanez, et al. (2001). "Weight change in old age and its association with mortality." Journal of the American Geriatrics Society **49**(10): 1309-1318.
Ogden, C. L., M. D. Carroll, et al. (2014). "Prevalence of childhood and adult obesity in the United States, 2011-2012." JAMA **311**(8): 806-814.
Oreopoulos, A., R. Padwal, et al. (2008). "Body mass index and mortality in heart failure: A meta-analysis." American Heart Journal **156**(1): 13-22.
Pinkston, J. C. (2014). The Dynamic Effects of Obesity on the Wages of Young Workers.
Rowland, M. L. (1990). "Self-reported weight and height." Am J Clin Nutr **52**(6): 1125-1133.
Sahyoun, N. R., L. M. Maynard, et al. (2008). "Factors associated with errors in self-reported height and weight in older adults." J Nutr Health Aging **12**(2): 108-115.
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." Public Health **9**: 421.

Thorpe, K. E. and D. H. Howard (2006). "The Rise In Spending Among Medicare Beneficiaries: The Role Of Chronic Disease Prevalence And Changes In Treatment Intensity." Health Affairs **25**(5): w378-w388.