The Journal of the Economics of Ageing xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

The Journal of the Economics of Ageing

journal homepage: www.elsevier.com/locate/jeoa

Full length article

Reporting error in weight and height among older adults: Implications for estimating healthcare costs

John Cawley [a,b], Johanna Catherine Maclean [c,*,1], Asia Sikora Kessler [d]

[a] Department of Policy Analysis and Management, Cornell University, NBER & IZA, Ithaca, NY, United States
[b] School of Economics, University of Sydney, Sydney, Australia
[c] Department of Economics, Temple University, NBER & IZA, Philadelphia, PA, United States
[d] Department of Health Promotion, Social and Economic Behavioral Health, University of Nebraska Medical Center, Omaha, NE, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Healthcare costs
Reporting error
Weight
Height
Obesity
Older adults

ABSTRACT

Previous research has identified obesity as a major contributor to healthcare costs among older adults. A limitation of this literature is its reliance on self-reported measures of weight and height, which may contain substantial error that can lead to bias in estimates of obesity prevalence and coefficients in healthcare utilization regressions.

This paper estimates the extent of reporting error in weight and height among older adults, and examines its implications for estimating healthcare costs within this population. Moreover, this paper is the first to apply methods to correct for reporting error in weight and height to an older adult sample, and examine the extent to which these methods reduce bias from reporting error. Previous research on reporting error in weight has focused on working-age adults, but older adults are likely to have different patterns of misreporting due to declines in cognition, changes in body composition, and other age-related factors.

We find substantial error in older adults' reports of weight and height, and this error is neither classical nor independent of common regressors in econometric models. Use of self-reports leads to bias in estimates of mean BMI, obesity prevalence, coefficients in healthcare utilization regressions, and obesity-attributable healthcare costs. Correction algorithms can reduce but not eliminate bias, and in certain cases they can actually worsen bias. These findings have implications for accurate estimates of the impact of obesity on healthcare utilization and costs among older adults.

Introduction

Understanding the determinants of healthcare costs among older adults (defined as 65 years and older) in the United States is important from a public finance perspective because older adults have direct access to public health insurance through Medicare. The costs of financing Medicare are substantial: $585.7 billion in 2013 (Centers for Medicare & Medicaid Services, 2014), which represents 20% of total healthcare expenditures (Centers for Medicare & Medicaid Services, 2014) and 14.5% of the Federal budget (Kaiser Family Foundation, 2014) in the U.S.

The financial sustainability of the Medicare program is of concern to policy makers (Baicker et al., 2013; Davis et al., 2013; Eibner et al., 2013). Such concern is driven by increases in healthcare costs generally (Hartman et al., 2014) and increases in healthcare utilization during later stages of life in particular (Alemayehu and Warner, 2004). The growing share of older adults in the population adds to this concern: those aged 65 and older represented 14.1% of the U.S. population in 2013 (authors' calculations using the 2013 American Community Survey), and are projected to rise to 16.6% by 2020 (Anderson and Hussey, 2000). The most recent estimates suggest that the trust fund for Medicare Part A, which covers hospitalizations, will be depleted by 2030 (Kaiser Family Foundation, 2014). Identifying costly health conditions and implementing cost-effective programs to prevent and reduce these conditions is a potential approach to sustain Medicare.

Obesity is a possible health condition to target as it is considered a major contributor to Medicare costs (Finkelstein et al., 2003, 2009; Lakdawalla et al., 2005; Thorpe and Howard, 2006; Yang and Hall, 2008). Obesity is prevalent among older adults; 35.4% of Americans aged 60 and older were obese in 2011–12 (Ogden

* Corresponding author.
E-mail addresses: johncawley@cornell.edu (J. Cawley), catherine.maclean@temple.edu (J.C. Maclean), asia.sikora@unmc.edu (A. Sikora Kessler).

[1] Authors listed in alphabetical order. All authors contributed equally to this manuscript.

http://dx.doi.org/10.1016/j.jeoa.2016.10.001
2212-828X/

et al., 2014), and obese beneficiaries cost Medicare $600 more per year than healthy weight beneficiaries (Finkelstein et al., 2009).

A necessary condition for careful study of obesity-attributable healthcare utilization and costs among older adults – or indeed any demographic group – is the ability to accurately measure obesity in data. Ideally, physical measurements of weight and height collected by medical professionals would be used, but nationally representative datasets containing healthcare measures (e.g., utilization) typically do not collect physical measurements, and instead rely on respondent reports of weight and height (e.g., the Medicare Current Beneficiary Survey [MCBS] and the Medical Expenditure Panel Survey [MEPS]). Thus, a limitation of previous economic studies of the healthcare costs of obesity among older Americans is that they have generally relied on self-reported weight and height, rather than physical measurements.

If weight and height are reported with error, coefficients in regressions of healthcare utilization and costs may be biased. Policy decisions based on biased estimates may lead to sub-optimal investment in obesity-related programs. For example, if the reporting error causes upward bias in estimates of the healthcare utilization and costs of older adult obesity, governments may over-invest in anti-obesity programs targeting older adults. Alternatively, if reporting error causes attenuation bias, it could result in underinvestment in obesity prevention and treatment. Accurate estimates are needed to provide guidance to Medicare policy makers as they devise strategies to address high and rising healthcare costs among older adults.

This paper studies reporting error in weight and height among older adults, and its implications for healthcare research. As such, it relates to a series of economic studies that examine the amount of, nature, and consequences of reporting error for the working age population (Cawley and Burkhauser, 2006; Cawley et al., 2015a; Courtemanche et al., 2015). The current study makes three important contributions to economic literature on reporting error in weight and height.

First, it provides new and up-to-date information on the extent and nature of self-report error in older adults, using the Health and Retirement Study (HRS), a dataset specifically designed to study older adult health and healthcare use that has been frequently used in the literature studying the implications of obesity in older adults (Michaud et al., 2007; Goldman et al., 2009; Mehta and Chang, 2009; Monteverde et al., 2010; Wilson, 2012). Second, we examine the impact of reporting error on estimates of mean BMI, the prevalence of obesity, regression coefficients in models of healthcare utilization, and obesity-attributable healthcare costs. Many researchers believe that self-reporting error biases regression coefficients toward zero (Bound et al., 2001). However, bias is attenuating only under certain conditions; we explicitly test these conditions using the HRS. Third, we are the first to apply methods to correct for reporting error in weight and height to an older adult sample, and examine the extent to which these methods reduce bias from reporting error.

This paper is organized as follows. The second section provides background on reporting error in survey data and a discussion of related research on reporting error. We discuss our data, variables, and methods in the third section. Results are reported in the fourth section and the last section concludes.

Background on reporting error

Bound et al. (2001) provide a comprehensive review of the literature on reporting error in survey data. The authors state that researchers tend to assume that reporting error in a regressor causes attenuation bias in estimates of regression coefficients. However, the authors demonstrate that this assumption is true only under certain conditions: 1) if the regression model is linear, 2) there is a single regressor in the model or the reporting error is uncorrelated with other regressors, and 3) the reporting error in the regressor is classical, i.e., uncorrelated with the true value (Bound et al., 2001).

In many cases, reporting error in weight may not satisfy these conditions. Error in binary variables such as weight classifications (e.g., obesity) *cannot* be classical because error is negatively correlated with the true value. Continuous variables such as weight in pounds (or kg) or body mass index are also unlikely to have classical reporting error because individuals to under-report their weight, a tendency that increases with true weight (Rowland, 1990), with the result that reporting error is correlated with the true value of the variable. Moreover, common regressors in economic models (e.g., education, sex, age) are potentially correlated with reporting error in weight and/or height. In addition, many healthcare outcomes (e.g., diagnoses, hospitalizations) are binary or counts and thus suggest the use of a non-linear model (e.g., probit, negative binomial). Bound et al. (2001) argue that the direction of the bias is difficult to sign *ex ante* in most applications and reliance on heuristics (e.g., reporting error always leads to attenuation bias in regression coefficient estimates) can lead researchers to incorrect interpretations of their findings.

A number of medical studies have investigated reporting error in weight and height among older adults (Gunnell et al., 2000; Kuczmarski et al., 2001; Gorber et al., 2007; Craig and Adams, 2009; Stommel and Schoenborn, 2009). These studies conclude that reports of weight and height by older adults contain substantial and systematic reporting error; e.g. weight is generally underreported and cognitive impairments are associated with less accurate reports.

We build on this research in three important ways. First, we provide estimates of the nature of reporting error in weight for a large, nationally representative sample of older adults from recent years (2006–2012). Second, we go beyond simply documenting the nature of reporting error and examine its consequences for bias in estimates of mean body mass index (BMI), the prevalence of obesity, and the association between obesity and healthcare utilization and costs. This additional information is necessary to understand the practical importance of the reporting error for economic studies.

Finally, we examine the extent to which algorithms developed to correct for reporting error using information from datasets that contain both measured and reported measurements (referred to as 'validation' datasets in the statistical literature on error-in-variables) address reporting error in datasets that contain only reports of height and weight ('primary' datasets). Economists have developed such correction methods to address error in weight and height reports based on the above-noted statistical literature on errors-in-variables to (Cawley, 2004; Cawley and Burkhauser, 2006; Courtemanche et al., 2015). However, to the best of our knowledge, these studies have focused exclusively on working age adults.

In their seminal study on validation data methods, Lee and Sepanski (1995) argue that in order for a dataset to be used as validation dataset for the correction of reporting error two assumptions must be satisfied. First, the distribution of the variable in the validation dataset must be the same as the distribution of the variable in the primary dataset (this is referred to a having a 'surrogate' variable). Second, the surrogate variable in the validation dataset must satisfy transportability across datasets, meaning that the distribution of the incorrectly measured variable conditional on the validated variable is the same in both datasets.[2]

[2] Lee and Sepanski (1995) rely on a weaker assumption in their study: the equivalence in the expected values of the variable (measured weight or height in our study) across the validation and primary datasets, conditional on the value of surrogate variable (reported weight our height in our study).

Please cite this article in press as: Cawley, J., et al.. The Journal of the Economics of Ageing (2016), http://dx.doi.org/10.1016/j.jeoa.2016.10.001