Economics and Human Biology 19 (2015) 27–44

Contents lists available at ScienceDirect

Economics and Human Biology

journal homepage: http://www.elsevier.com/locate/ehb

Reporting error in weight and its implications for bias in economic models

John Cawley [a,b,*], Johanna Catherine Maclean [c,1], Mette Hammer [d], Neil Wintfeld [d]

[a] Department of Policy Analysis and Management, Cornell University, 2312 MVR Hall, Ithaca, NY 14853, United States
[b] School of Economics, University of Sydney, Sydney, NSW, Australia
[c] Department of Economics, Temple University, Ritter Annex 869-1301, Cecil B Moore Avenue, Philadelphia, PA 19122, United States
[d] Novo Nordisk, Inc., 800 Scudders Mill Road, Plainsboro, NJ 08536, United States

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form 13 July 2015
Accepted 14 July 2015
Available online 23 July 2015

Keywords:
Weight
Obesity
Reporting error
Bias

ABSTRACT

Most research on the economic consequences of obesity uses data on self-reported weight, which contains reporting error that has the potential to bias coefficient estimates in economic models. The purpose of this paper is to measure the extent and characteristics of reporting error in weight, and to examine its impact on regression coefficients in models of the healthcare consequences of obesity.

We analyze data from the National Health and Nutrition Examination Survey (NHANES) for 2003–2010, which includes both self-reports and measurements of weight and height. We find that reporting error in weight is non-classical: underweight respondents tend to overreport, and overweight and obese respondents tend to underreport, their weight, with underreporting increasing in measured weight. This error results in roughly 1 out of 7 obese individuals being misclassified as non-obese. Reporting error is also correlated with other common regressors in economic models, such as education.

Although it is a common misconception that reporting error always causes attenuation bias, comparisons of models that use self-reported and measured weight confirm that reporting error can cause upward bias in coefficient estimates. For example, use of self-reports leads to overestimates of the probability that an obese man uses a prescription drug, has a healthcare visit, or has a hospital admission.

These findings underscore that models of the consequences of obesity should use measurements of weight, when available, and that social science datasets should measure weight rather than simply ask subjects to report their weight.

* Corresponding author at: Department of Policy Analysis and Management, Cornell University, 2312 MVR Hall, Ithaca, NY 14853, United States. Tel.: +1 607 255 0952; fax: +1 607 255 4071.

E-mail addresses: JHC38@cornell.edu (J. Cawley), catherine.maclean@temple.edu (J.C. Maclean), mthm@novonordisk.com (M. Hammer), nwif@novonordisk.com (N. Wintfeld).

[1] Tel.: +1 215 204 0560.

http://dx.doi.org/10.1016/j.ehb.2015.07.001
1570-677X/

1. Introduction

The prevalence of obesity in the United States has more than doubled since 1980 (Burkhauser et al., 2009)[2]. As of 2011–2012, 34.9% of adults in the U.S. are obese (Ogden et al., 2014), and the Surgeon General has declared that obesity in the U.S. has reached epidemic proportions (U.S. DHHS, 2010).

This has led to a substantial amount of research on the economic consequences of obesity, such as healthcare costs, job absenteeism, wages, and employment (for reviews, see Cawley, forthcoming; Averett, 2011; Finkelstein and Yang, 2011). Many of these studies use self-reported weight and height, because only self-reported, and not measured, values are available in most social science datasets. For example, the Medical Expenditure Panel Survey (MEPS), the primary dataset used to estimate the healthcare costs of obesity, and the National Longitudinal Survey of Youth (NLSY), the primary dataset used to estimate the impact of obesity on wages, both include self-reports but not measurements of weight. Reporting error has the potential to bias coefficient estimates (Bound et al., 2001), which in this context can result in misleading estimates of the economic consequences of obesity.

A comprehensive review of the consequences of error in economic variables, Bound et al. (2001), states that many 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 coefficient estimate is biased downward (attenuated). Bound et al. (2001) show that, in fact, 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 in the model. For a detailed explanation of the determinants of the sign of the bias, we refer readers to Bound et al. (2001), but the following simple model derived from that work is helpful in summarizing the basic implications of reporting error for regression coefficients.

Assume that the true model is: $y^* = X^* \beta + \varepsilon$ where y^* and ε are scalars and X^* and β are vectors. Suppose that the researchers do not observe the true values y^* and X^*, only the reported values that are characterized by reporting error:

$$X = X* + u$$
$$y = y* + v$$

If there is only one independent variable in the regression model, and it is characterized by classical measurement error (defined as u and v being uncorrelated with X^*, y^*, and ε), then the coefficient suffers from attenuation bias equal to the noise-to-signal ratio in the mismeasured regressor:

$$\hat{\beta} = \beta\left[1 - \frac{\sigma_u^2}{\sigma_{X*}^2 + \sigma_u^2}\right]$$

Reporting error in a continuous variable such as weight or BMI could be classical, but error in binary variables such as clinical weight classifications (e.g. obesity) cannot be classical because error is negatively correlated with the true value of the variable.

If the error u violates the classical assumption and is correlated with the true X^*, then it is possible for reporting error to cause either upward or downward bias in $\hat{\beta}$ (upward bias is possible if the error is mean reverting; i.e. u and X^* are negatively correlated). If the regression includes more than one regressor, then the bias in the coefficient on X^* depends in part on the collinearity between the error-ridden variable X^* and the other regressors. In this case, the bias on any given coefficient estimate is difficult to characterize; even if the error is classical, the bias could be in either direction. In summary, only if there is a single regressor and its error is classical is the bias ensured to be attenuating; if the error is not classical or there are multiple regressors that are measured with error (irrespective of whether the error is classical) then the bias could be in either direction. This framework suggests that (1) relying on simple heuristics can lead researchers to inaccurate conclusions; and (2) signing the bias attributable to reporting error requires strong assumptions concerning a number of factors that are difficult to test.

So far this discussion has assumed a linear model. Bound et al. (2001) report that results for non-linear (e.g. probit, negative binomial) models are generally comparable to those for linear models, but if anything, nonlinearities tend to worsen the biases associated with measurement error.

This paper answers questions that arise from this discussion, specifically: whether reporting error in weight, height, and the implied error in body mass index (BMI) is mean zero, whether the error is classical or is correlated with measured weight, and whether the reporting error in weight is correlated with other commonly-utilized regressors in models of the economic consequences of obesity. We also examine how use of self-reports rather than measurements biases regression coefficients in models of the association of obesity with healthcare utilization, specifically: healthcare visits, hospital admissions, and prescription drug utilization. As an extension, we illustrate the magnitude of the bias by showing its implications for the estimated hospitalization costs associated with obesity.

The findings of this paper are relevant for the substantial literature that uses self-reports of weight to estimate the economic consequences of obesity (see the reviews in: Cawley, forthcoming; Averett, 2011; Finkelstein and Yang, 2011). Specifically, use of self-reported values may lead to inaccurate estimates of the healthcare costs of obesity, which could then lead to suboptimal investment in prevention and treatment programs.

This paper also relates to the general literature on reporting error in weight. A number of studies have documented reporting error in weight (see, e.g. Gil and Mora, 2011; Danubio et al., 2008; Ossiander et al., 2004). A comprehensive review concludes that there is a tendency for weight to be underreported (Connor Gorber et al., 2007); however, the reviewed studies differed in quality and no overall effect size could be estimated. Four studies

[2] For adults, obesity is defined as a body mass index (or BMI, which is calculated as weight in kg divided by height in m squared) of 30 or higher (see U.S. DHHS, 2010).