IZA Journal of Labor Economics
a SpringerOpen Journal

ORIGINAL ARTICLE   **Open Access**

The consequences of measurement error when estimating the impact of obesity on income

Donal O'Neill[*] and Olive Sweetman

* Correspondence:
donal.oneill@nuim.ie
Department of Economics, Finance
and Accounting, National University
of Ireland Maynooth, Maynooth, Co,
Kildare, Ireland

Abstract

This paper examines the consequences of using self-reported measures of BMI when estimating the effect of BMI on income for women using both Irish and US data. We find that self-reported BMI is subject to substantial measurement error and that this error deviates from classical measurement error. These errors cause the traditional least squares estimator to overestimate the relationship between BMI and income. We show that neither the conditional expectation estimator nor the instrumental variables approach adequately address the bias and briefly discuss alternative approaches that could be considered when faced with non-classical measurement error.

JEL codes: C13, C26, I14

Keywords: Obesity, Non-classical measurement error, Auxiliary data, Instrumental variables

1. Introduction

Obesity is a medical condition described as excess body weight in the form of fat. The International Association for the Study of Obesity (IASO) reports that approximately 1.5 billion adults are currently overweight and of these, 525 million are obese (IASCO 2010). Obesity is an important cause of morbidity, disability and premature death and increases the risk for a wide range of chronic diseases (WHO 2009). As a result there are substantial direct and indirect costs associated with obesity that put a strain on a country's resources. There have also been a number of studies that examine the impact of obesity on individual outcomes such as wages (Cawley 2004; Brunello and d'Hombres 2007), labour force participation and employment (De Sousa 2012) and educational achievement (Kaestner and Grossman 2009; von Hinke et al. 2012). Many of these studies find a significant negative association between body weight and individual success, so that the costs of obesity are borne at the individual as well as the national level. This is especially true for women who are the focus of this paper.

The most widely-used measure of obesity is an individual's Body Mass Index (BMI), defined as weight in kg/height in m^2.[1] The majority of studies using BMI rely on self-reported measures from survey data sets such as the National Longitudinal Study of Youth, the European Community Household Panel and the National Child Development Study. However, there is a large body of evidence that suggests that self-reported BMI tends to underestimate true BMI; this occurs both because people underreport their weight and overstate their height (see for example Conor Gorber et al. 2007).

Most authors recognize this and as result typically adopt a range of approaches to deal with the problem of mismeasured obesity.

In this paper we use two different data sets to examine the nature of measurement error in self-reported female BMI and to illustrate the consequences of these errors. The Growing up in Ireland Survey (GUI) and the National Health and Nutrition Examination Survey (NHANES) both contain self-reported and recorded measures of height and weight. The availability of both self-reported and recorded measures allows us to examine the nature of measurement error in BMI in detail, while carrying out the analysis on two independent data sets helps establish general findings that are unlikely to be specific to a particular setting or country.

In keeping with previous work we show that measurement error in self-reported BMI is substantial and correlated with the true recorded value of BMI. We extend previous work by showing that, for the outcomes we examine, the self-reported error is also differential. This means that the reported BMI contains information on the respondent, even after controlling for true BMI. We show that the bias resulting from these errors can be substantial and that popular approaches proposed in previous work may not adequately address these problems. In the conclusion we discuss some alternative approaches that could be considered to reduce the bias.

The layout of the paper is as follows. Section 2 summarises the statistical literature on measurement error. We focus on the biases that arise when the assumptions of classical measurement error are relaxed. Section 3, discusses our data and examines the nature of measurement error in the self-reported measures of BMI. Section 4 considers the implications of this measurement error when examining the impact of obesity on income. Section 5 concludes and discusses the implications of our findings.

2. Measurement error in economic analysis

It is well known that measurement error in observed data can lead researchers to draw incorrect inferences (Fuller 1987; Carroll et al. 1994; Bound et al. 2001; Hyslop and Imbens 2001). Much of the early work in this area focused on the typical textbook model of classical measurement error. However in their study of measurement error in labour market data, Bound et al. (1994) suggested that the assumption of classical measurement error was often a matter of convenience rather than conviction. Bound et al. (2001), pg.3709 conclude their survey of measurement error by calling for researchers to pay greater attention to the possibility of non-classical measurement error, both in assessing the likely biases in the analyses that take no account of measurement error and in devising procedures that correct for such error.

In recent years a number of papers have examined the consequence of non-classical measurement error in labour economics. Pischke (1995), O'Neill et al. (2007) and Gottschalk and Huynh (2010) all show that non-classical measurement error, of the type typically found in income data, attenuates the role of white noise measurement error in models of earnings dynamics, while Kim and Solon (2005) suggest that real-wages may be even more procyclical than recent studies suggest, once one accounts for mean-reverting measurement error. Looking at measurement error in self-reported BMI specifically, Plankey et al. (1997) examine the consequences of measurement error in self-reported BMI when classifying people according to obesity status but do not

consider the consequences of measurement error for regression analysis. Bauhoff (2011) analysed the impact of measurement error in self-reported BMI when BMI is used as the dependent variable in a treatment model. However, he did not consider situations where BMI is an explanatory variable, nor did he consider alternative estimation procedures. Stommel and Schoenborn (2009) compare self-reported and recorded BMI using US data and find a substantial amount of misclassification of obesity status when using self-reported BMI, particularly in the extreme (over – or underweight) categories. They examine the consequences of this measurement error when analysing the impact of BMI on a range of health risks. However, we are not aware of any previous study that has examined the consequences of measurement in self-reported BMI when estimating the relationship between obesity and labour market outcomes, or of any studies that have examined the consequences of measurement error for the variety of estimators we consider in this paper.

Bound et al. (2001) summarise a number of the approaches that have been suggested for dealing with measurement error. One popular approach relies on the availability of auxiliary data. While auxiliary data allow researchers to examine the nature of measurement error, these data typically do not contain information on the dependent variable of interest. As a result the information gained from the auxiliary data must be "transported" into the main survey data. One approach is to use the auxiliary data to estimate the expected value of the true variable conditional on the reported value and then to use this model to estimate expected BMI in the original survey data. This approach is known as the "conditional expectation (CE)" approach (Lyles and Kupper 1997) or the "regression calibration" approach (Guo and Little 2011) and has been used to correct for measurement error in previous studies looking at the impact of obesity on labour market outcomes (Cawley 2000, 2004; Lindeboom et al. 2010). A second approach uses the instrumental variable (IV) estimator to obtain consistent estimates in the presence of measurement error.[2] With classical measurement error, both of these approaches provide consistent estimates but it is easy to show that these standard approaches do not work if measurement error is non-classical.

The purpose of our paper is not to estimate a true overall effect of obesity on income, accounting for all possible problems that might arise in doing so. Rather, our objective is more straightforward; namely to demonstrate the bias that arises when using self-reported BMI instead of true BMI in labour market analysis and to evaluate a number of alternative approaches that have been suggested for tackling the problem of measurement error in practice. In order to focus attention on measurement error, we consider estimating a conditional mean function, $E[Y|X^*]$, where Y is the outcome of interest (assumed measured without error) and X^* is the true value of the variable of interest.[3] Assuming $E[Y|X^*]=\beta X^*$ we write our regression equation as:

$$Y = \beta X^* + \epsilon. \qquad (1)$$

By construction $E[\epsilon X^*] = 0$. This allows us to abstract from potential problems associated with endogeneity of X^*. We will briefly return to this issue later. In addition the observed value of X^*, which we denote by X, is given by

$$X = X^* + u \qquad (2)$$

where u is the measurement error. Classical measurement error typically refers to the

situation where the error for any individual, u_i, is unrelated to the true value X_i^*; this in turn implies $E(X^*u) = 0$. Non-classical measurement error can arise in two cases; firstly there may be a relationship between the reported measurement error and the true value so that $E(X^*u) \neq 0$; secondly there may be a relationship between the reported measurement error and the residual in equation (1) so that $E(\epsilon u) \neq 0$. The latter situation is sometimes referred to as differential measurement error; in this case X contains information about Y even after we condition on X^*.[4]

The probability limit of the OLS estimator from the regression of Y on the observed X is given by:

$$plim \ b_{OLS} = \frac{\beta\{Var(X^*) + Cov(u, X^*)\}}{var(X^*) + var(u) + 2Cov(u, X^*)} + \frac{Cov(u, \varepsilon)}{var(X^*) + var(u) + 2Cov(u, X^*)} \tag{3}$$

If $Cov(u, X^*) = Cov(u, \varepsilon) = 0$ this simplifies to the textbook attenuation bias associated with classical measurement error. However, it is well known that violation of either of these conditions will alter the probability limit of the OLS estimator such that the direction of the inconsistency cannot be established a-priori.

The problems posed by non-classical measurement error for both the CE and IV approach are also immediate. The CE approach regresses the true value, X^*, on the self-reported measure, X, using auxiliary data and uses these coefficients to predict $E[X^*|X]$ in the survey data. To see how this approach works consider taking conditional expectations in equation (1):

$$E[Y|X = x] = \beta E[X^*|X = x] + E[\epsilon|X = x] \tag{4}$$

$$= \beta E[X^*|X = x] + E[\epsilon|X^* + u = x] \tag{5}$$

$$= \beta E[X^*|X] + E_{X^*}[E[\epsilon|X^* + u = x, X^* = x^*]|X = x] \tag{6}$$

$$= \beta E[X^*|X] + E_{X^*}[E[\epsilon|u = x - x^*]|X = x] \tag{7}$$

Clearly if $E[\epsilon|u] = 0$ then a regression of Y on $E[X^*|X]$ will consistently estimate β. However, if $E[\epsilon|u] \neq 0$ a regression of Y on $E[X^*|X]$ alone will result in biased and inconsistent estimates of β.[5]

We can derive an expression for the probability limit of the CE estimator in cases where $E[X^*|X]$ is a linear function of X. In this case the CE estimator can be written as

$$b_{CE} = \left(\hat{X}^{*\prime}\hat{X}^*\right)^{-1}\left(\hat{X}^{*\prime}Y_{NV}\right),$$

where

$$\hat{X}^* = X_{NV}\left(X_V'X_V\right)^{-1}\left(X_V'X_V^*\right)$$

is the estimated value of $E[X^*|X]$ obtained by combining the validation sample (v) and the non-validation sample (nv).

In this case the CE estimator can be rewritten as

$$b_{CE} = \left(X_V' X_V^*\right)^{-1} C\left(X_{NV}' Y_{NV}\right).$$

where $C = \left(X_V' X_V\right)\left(X_{NV}' X_{NV}\right)^{-1}$ is a matrix that will converge in probability to the identity matrix under reasonable assumptions on the data generating processes in the validation and non-validation samples (see for example Angrist and Krueger 1992).

Written this way the CE estimator can be thought of as a two sample two stage least squares estimator in which X is used to instrument for X^* (Inoue and Solon 2010). In this case it is easy to show that the CE estimator converges to

$$plim\ b_{CE} = \frac{\beta[var(X^*) + cov(X^*, u)]}{[var(X^*) + cov(X^*, u)]} + \frac{cov(u, \varepsilon)}{var(X^*) + cov(X^*, u)} \tag{8}$$

$$= \beta + \frac{cov(u, \varepsilon)}{var(X^*) + cov(X^*, u)} \tag{9}$$

Furthermore, we can use these probability limits to compare the relative bias in the OLS and CE estimators. The following proposition follows immediately from a comparison of equations (3) and (9).

Proposition 1:

(a) $|plim\ b_{OLS}| = |plim\ b_{CE}| > |\beta|$
   if $-cov(X^*, u) = var(u)$ and $sgn(\beta) = sgn(cov(\varepsilon, u))$
(b) $|plim\ b_{OLS}| > |plim\ b_{CE}| > |\beta|$
   if $-cov(X^*, u) > var(u)$ and $sgn(\beta) = sgn(cov(\varepsilon, u))$

From equation (3) we see that the condition $-cov(X^*, u) = var(u)$ eliminates any bias in the OLS estimator resulting from correlation between X^* and the measurement error. When this condition holds the OLS estimator is similar to the CE estimator in that the only bias that arises is from the differential measurement error. A comparison of the denominator of the final terms in equations (3) and (9) also shows that this condition is sufficient to equalise the differential bias in the two estimators. When non-classical measurement error is relatively large, that is $-cov(X^*, u) > var(u)$, the correlation between X^* and the measurement error will cause a bias in the OLS estimator that is not present in the CE approach. In addition the bias due to differential measurement error will be larger in the OLS estimator than in the CE estimator. Both these effects will lead to the bias in the OLS estimator exceeding that in the CE approach.

In contrast to the CE approach which uses X to instrument X^*, the standard IV approach instruments the self-reported measure X with an instrument Z. The instrument Z should be such that $E(ZX^*) \neq 0$ and $E(Zu) = 0$. However, with non-classical measurement error, the correlation between X^* and u will mean that instruments that are strongly correlated with X^* are also likely to exhibit a correlation with the measurement error, thus violating the condition for consistency.[6] Formally the IV estimator converges to

$$plim\ b_{IV} = \frac{cov(Y,Z)}{cov(X,Z)} = \frac{cov(\beta X^* + \varepsilon, Z)}{cov(X^* + u, Z)} = \frac{\beta cov(X^*, Z)}{cov(X^*, Z) + cov(u, Z)} \qquad (10)$$

Clearly the second term in the denominator of this expression will lead to inconsistencies in the IV estimator; if this term is negative then the IV estimator may overestimate the true parameter of interest, β.

In the remainder of the paper we use data from two independent surveys to examine the nature of measurement error in self-reported BMI and to analyse the direction and magnitude of the potential biases discussed above in practice.

3. Data

Our analysis uses two data sets; the Growing Up in Ireland Survey (GUI) and the National Health and Nutrition Examination Survey (NHANES). The GUI data tracks the development of a cohort of Irish children born between November 1997 and October 1998. The data used for our analysis are from the first wave of interviews, which were carried out between August 2007 and May 2008. To our knowledge no one has previously used the GUI sample to examine the issue of measurement error in self-reported BMI. The NHANES III is a nationally representative survey of 33,994 individuals in the US aged two months of age and older. The interviews were carried out over the period from 1988–1994. The NHANES data have been used in previous studies looking at the impact of obesity on labour market outcomes (e.g. Cawley 2004). In this section we build on earlier studies that have used the NHANES to examine measurement error in self-reported BMI (e.g. Plankey et al. 1997; Burkhauser and Cawley 2008; Stommel and Schoenborn 2009).

The key feature of both these data sets is that, in addition to self-reported measures of height and weight, we also have independent measures of the respondent's height and weight. We refer to the latter as recorded measures and treat them as the true height and weight of the respondents. In the GUI sample the recorded measures were obtained by the interviewer in the respondent's home at the end of the interview. The respondent was unaware that these measurements would be taken at the time they were providing their self-reported measures. In the NHANES sample the health measurements were performed in specially-designed and equipped mobile centres, by a team of physicians, medical and health technicians, as well as dietary and health interviewers. These physical exams took place some weeks after the initial interview, and there is an expectation that respondents would have been aware of the physical exam at the time of the interview. In both samples we compare the recorded measures to the self-reported measures to determine the extent and the nature of measurement error in BMI. In addition both data sets contain information on household income which we use to consider the impact of measurement error on the estimated relationship between BMI and income. Since the incremental costs of obesity tend to be significantly higher for women than for men (Dor et al. 2010), we focus on the impact of obesity on economic outcomes for females to illustrate our findings.

For our purposes the GUI and NHANES data both have particular strengths and weaknesses. The GUI data has the advantage of relatively large samples. When we restrict the GUI sample to biological mothers of the study child who were not pregnant at the time of

the study we are left with a working sample size of 6637. When analysing the NHANES data we consider all white females between the ages of 18 and 65, who were not pregnant at the time of the study. Despite this broader age range the base sample size in NHANES is only 2332. In addition, and in contrast to the NHANES data where there was a lag of weeks between the self-reported and recorded measures, the self-reported and recorded BMI measures in the GUI were obtained within minutes of each other.

While the GUI benefits from the timing of data collection and the larger samples, the NHANES sample is representative of all women aged 18–65 in the US. In contrast the GUI sample is restricted to families who had a child aged nine at the time of the survey. The difference in the underlying populations is reflected in the age distributions in the two samples. Although the average age of the women in the GUI (39.89) is similar to that in the NHANES sample (41.35), the age variation in the Irish sample is much lower than in the US sample; the standard deviation of age in the GUI sample is 5.3 years compared to 13.59 years in the NHANES data.

The differences in timing of data collection, the sample sizes and the underlying populations between the two data sources suggest that neither of these data sets should be considered preferable, a-priori. However, conducting the analysis on two independent data sets from two different countries will be important in establishing general findings that are not specific to one particular setting.
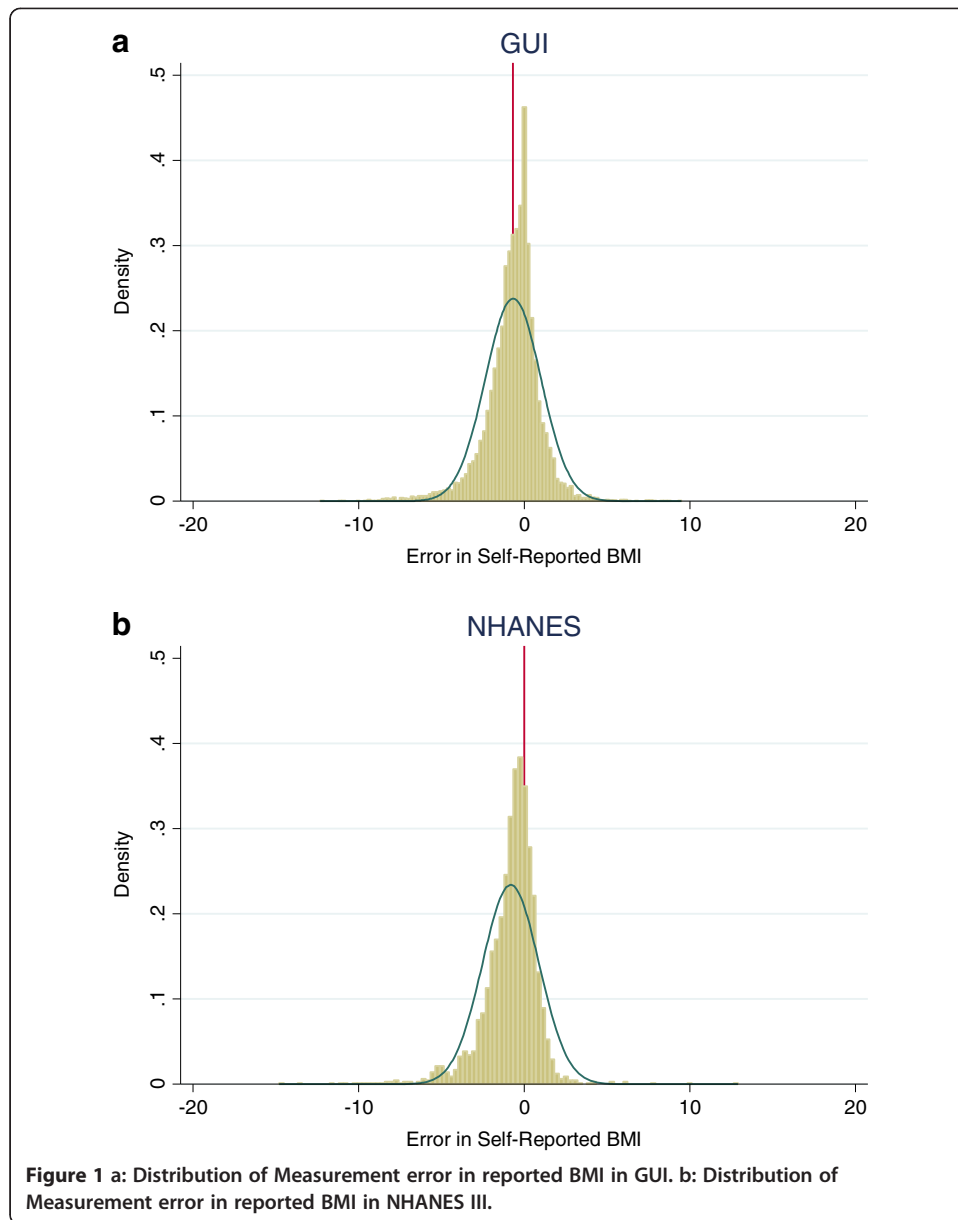
Summary statistics for the self-reported and recorded measures of BMI are given in Table 1. The top panel reports the results for GUI and the bottom panel for NHANES.[7] Using these data we find that 42.65% (13.9%) of the mothers in the GUI sample are overweight (obese) on the basis of self-reported data. However, the true numbers are 49.55% and 17.34%. The corresponding figures for the NHANES data are 42.075% (17.80%) for the self-reported data compared to true rates of 46.57% and 22.86%. The tendency for respondents in our sample to underestimate their BMI in self-reported data is evident in both these data sets and is consistent with previous findings.[8]

To examine the extent of measurement error in the self-reported measures of BMI, we calculate the error in the self-reported data by subtracting the true BMI measure from self-reported measure. The density of the measurement error is given in Figure 1a and 1b, while summary statistics for the measurement error are given in Table 2. The results in Table 2 show that the mean error is negative in both samples. In addition the magnitude of the measurement error is sizeable; the standard deviation in the measurement error accounts for 34.1% of the standard deviation of the true BMI variable in the Irish data and 27% in the US data.

Since an individual's BMI depends on both their height and weight it is useful to consider these measures separately in order to determine their contribution to

Table 1 Summary statistics on recorded and self-reported BMI

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
		GUI			
Recorded BMI	6637	25.99	4.92	15.06	64.16
Self-reported BMI	6637	25.31	4.61	12.70	69.44
		NHANES			
Recorded BMI	2332	26.12	6.33	15.1	68.5
Self-reported BMI	2332	25.31	5.77	10.9	72.6

Figure 1 a: Distribution of Measurement error in reported BMI in GUI. b: Distribution of Measurement error in reported BMI in NHANES III.

measurement error in overall BMI. Tables 3 and 4 report the summary statistics for self-reported and recorded height and weight respectively. These data show that while height and weight are both measured with error in Ireland and the US, misreporting of weight is the dominant factor in accounting for measurement error in BMI.

The remainder of the paper explores the discrepancies between true and reported BMI in more detail and considers the implications of these differences for the estimated relationship between BMI and income.

4. Analysis and results

In the previous section we showed that females in both the GUI survey and the NHANES III data are more likely to underreport their BMI. In the notation of section 2,

Table 2 Summary statistics on self-reported measurement error in BMI

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
		GUI			
Self-reported BMI error	6637	−.672	1.68	−12.33	9.50
		NHANES			
Self-reported BMI error	2332	−.8106	1.71	−14.79	12.92

this implies $E[u] < 0$. However, as noted in section 2, the implications of measurement error for economic analysis will differ depending on whether the error is classical or non-classical in nature. To determine this we examine the relationship between the error and the true measure of BMI. Figures 2a and 2b graph the relationship between the measurement error and the true measure of BMI in both data sets. From this we see a negative relationship between the level of measurement error and the true value of BMI in both samples. People with higher BMI's are more likely to underreport their BMI.[9] The correlation is −.35 in the Irish sample and −.45 in the US data. This negative correlation illustrates the non-classical nature of measurement error in BMI in our data and will be important in determining the biases associated with each of the estimators we consider.

To examine the consequences of non-classical measurement error in BMI we consider the relationship between BMI and income. Ideally we would like to look at the relationship between individual income and BMI. However individual income data are not available in either of our data sets. Instead we are restricted to examining the relationship between a woman's BMI and her total household income (see also Parks et al. 2011). While this is somewhat different to individual wages typically examined in previous work, the use of household income as the dependent variable nevertheless provides a useful framework for illustrating the consequences of non-classical measurement error.[10]

4.1 BMI and household income (GUI)

To begin, we estimate the relationship between true BMI and the level of household income in the GUI.[11] Apart from sampling error this coefficient gives the "true" relationship between BMI and income. This establishes the benchmark for the true parameter of interest in our analysis.[12] The results are given in the first column of Table 5. The estimated coefficient on true BMI shows a significant negative relationship between BMI

Table 3 Summary statistics on recorded and self-reported height (cms)

Variable	Observations	Mean	St. Deviation	Minimum	Maximum
		GUI			
Recorded height	6637	163.21	6.19	139	189
Self-reported height	6637	163.72	6.55	137	189
Correlation between error and recorded height			−0.10		
		NHANES			
Recorded height	2332	163.24	6.31	137	183
Self-reported height	2332	163.35	6.63	122	183
Correlation between error and recorded height			−0.062		

Table 4 Summary statistics on recorded and self-reported weight (Kgs)

Variable	Observations	Mean	St. Deviation	Minimum	Maximum
	GUI				
Recorded weight	6637	69.19	13.46	36	177
Self-reported weight	6637	67.87	12.65	30	178
Correlation between error and recorded weight			−0.35		
	NHANES				
Recorded weight	2332	69.57	17.10	37.45	175.45
Self-reported weight	2332	67.54	15.83	29.70	180.00
Correlation between error and recorded weight			−0.43		

and income. The true parameter estimate is −.821, which indicates that a 5 point increase in BMI (which corresponds to approximately a 1 standard deviation increase) is associated with a €4105 reduction in income.

To examine the impact of measurement error we estimate the same regression only this time using self-reported BMI. The results are given in the second column of Table 5. In contrast to what we would expect with classical measurement error, we see that using the self-reported BMI *overstates* rather than attenuates the relationship between BMI and income. A statistical test of the hypothesis that the coefficients on the recorded BMI and self-reported BMI are equal is rejected with a p-value of .0008.[13] Using self-reported BMI we estimate that a 5 point increase in BMI results in a €4605 reduction in income. As a result the use of self-reported BMI overstates the loss of income by approximately 12.18%.

We next consider the extent to which the alternative estimators proposed in the literature succeed in tackling the bias associated with self-reported BMI. We first consider the use of the CE estimator (e.g. Cawley 2004). Following the CE approach we first regress true BMI on self-reported BMI.[14] The coefficients from this first stage regression are then used to adjust the self-reported data. The predicted measure is then used in place of the self-reported measure in the regression analysis.

The results from the second stage are presented in the final column of Table 5. The results show that the CE approach remains biased. In this application the bias in the CE approach is almost identical to that with the original self-reported data. In section 2 we showed that the bias in the OLS and CE estimator is the same whenever $-cov(X^*, u) = var(u)$. In our application $-cov(X^*, u) = 2.88$ and $var(u) = 2.81$, which explains the similarity of the estimates from the two approaches. Both estimators are inconsistent, with the magnitude of the inconsistency being determined by the nonclassical and nondifferential nature of the measurement error. Finding that the OLS and CE estimator may be very similar, and yet both inconsistent, serves as a warning against using the similarity of corrected and uncorrected estimates to infer the absence of measurement error bias.

The availability of internal validation in our study allows us to examine the issue of differential measurement error in more detail. We can obtain consistent estimates of ε in equation (1) from the residuals of a regression of income on recorded BMI. Under the assumption of non-differential measurement error these residuals should be uncorrelated with the observed measurement error u. To examine this we regress the predicted residuals on the measurement error. The coefficient on the measurement
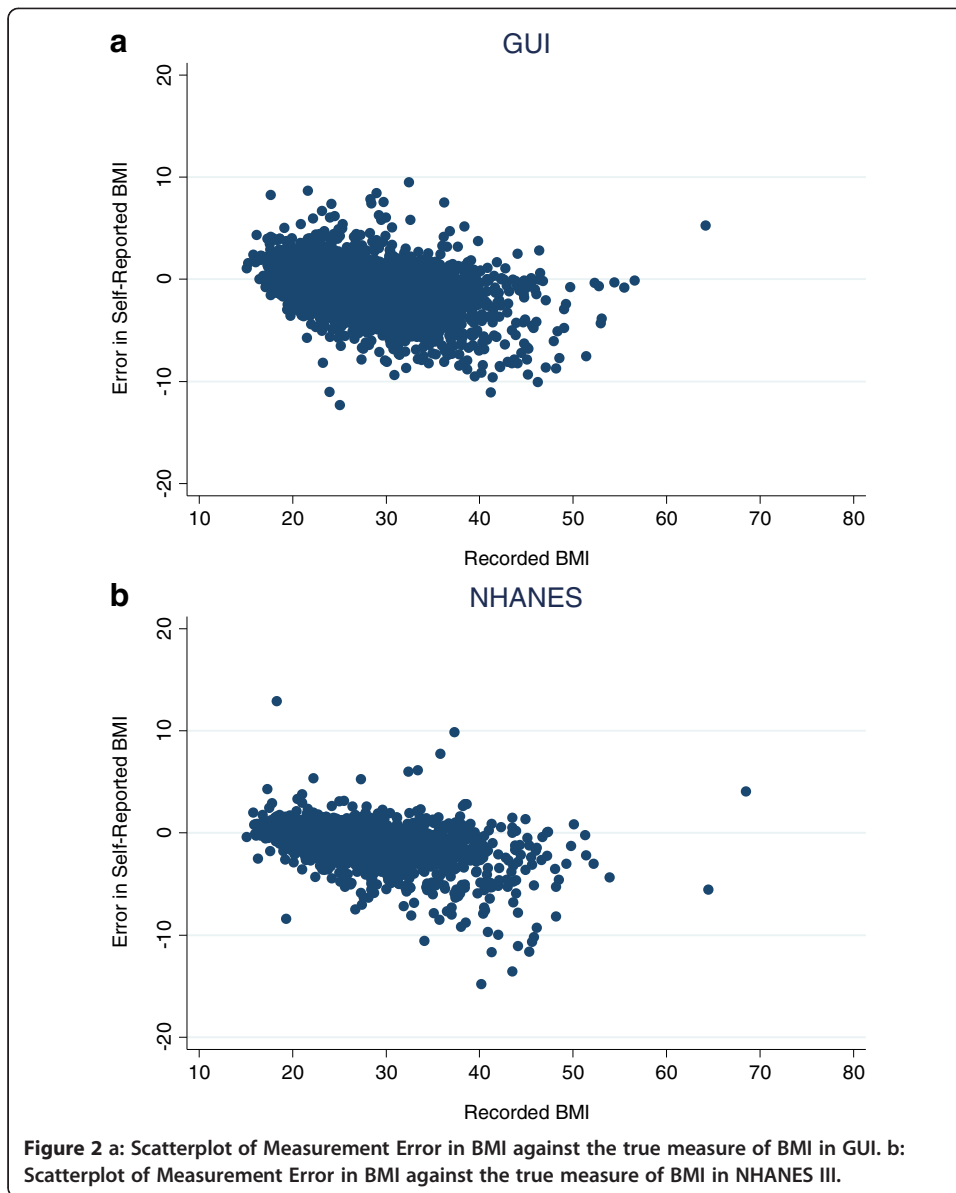
Figure 2 a: Scatterplot of Measurement Error in BMI against the true measure of BMI in GUI. **b:** Scatterplot of Measurement Error in BMI against the true measure of BMI in NHANES III.

Table 5 Estimated Coefficients on BMI in income regressions (GUI)

Dependent variable household income	OLS	OLS	CE
Recorded BMI	−.821		
	(.091)		
Self-reported BMI		−.921	
		(.097)	
Corrected BMI			−.918
			(.097)
N	6637	6637	6637

Income is measured in 000's; Standard error in brackets.

error is statistically significantly negative, with a point estimate of −0.74; the null-hypothesis of no relationship is rejected with a p-value = .006. It is this differential error that biases the estimates obtained using the CE approach.

It is possible that the problem of differential measurement error may be reduced by including additional variables as controls in the regression model. To consider this we re-estimate the income equation using a set of additional regressors to capture the range of controls typically used when estimating income equations in labour economics. These include controls for age, education, health status and marital status, a control for whether the respondent smokes on a daily basis and a control for English language proficiency.[15]

The results for this extended specification are given in Table 6. The first column reports the OLS results using recorded BMI, the second column gives the OLS results using self-reported BMI and the third column gives the results using the CE approach. Looking at the additional controls in the first column we see that they are all significant and have the expected signs. College educated, older, married women who are proficient in English receive an income premium, while women in poor health and those who smoke regularly have lower incomes. The coefficient on recorded BMI is negative and significant, though somewhat smaller in magnitude than the earlier estimate. This is to be expected since some of the effect of BMI on income may be operating through the health and education channels which are accounted for in the extended specification. The second column reports the results for self-reported BMI. The use of self-reported BMI has very little effect on the estimates for the additional controls. However, as before the use of self-reported BMI leads us to overestimate the impact of BMI on income. Again

Table 6 Estimated coefficients on BMI in income regressions obtained using CE approach with additional controls (GUI)

Dependent variable household income	OLS	OLS	CE
Recorded BMI	−.60		
	(.09)		
Self-reported BMI		−.67	
		(.10)	
Corrected BMI			−.67
			(.098)
College degree	21.63	21.58	21.58
	(1.02)	(1.03)	(1.03)
Age	.46	.46	.46
	(.096)	(.096)	(.096)
English language	12.62	12.71	12.71
	(2.03)	(2.02)	(2.02)
Married	21.23	21.28	21.28
	(1.70)	(1.69)	(1.69)
Poor health	−3.79	−3.67	−3.67
	(1.36)	(1.36)	(1.36)
Daily smoker	−5.97	−5.91	−5.91
	(1.25)	(1.25)	(1.25)
N	5965	5965	5965

Income is measured in 000's; Standard Error in brackets.

the bias is of the order of 12%. The results reported in the final column shows that even in this extended specification the CE estimate of the impact of BMI is biased and the bias is almost identical to that obtained using the original self-reported data.[16] It is clear from this that our earlier findings are robust and still evident even when we include a richer set of control variables in our income equation.

IV estimation is another popular approach used to overcome the bias associated with measurement error. A common instrument in this type of analysis is to use the BMI of a sibling or other relative to instrument for the respondent's BMI on the assumption that this should pick up genetic and environmental factors but may be unrelated to the measurement error (see Cawley 2000; Cawley et al. 2004; Brunello and D'Hombres 2007; Kline and Tobias 2008; Kortt and Leigh 2010; Lindeboom et al. 2010; Cawley and Meyerhoefer 2012). We follow Cawley (2000), Davey Smith et al. (2009), and Cawley and Meyerhoefer (2012) and use the recorded weight of a biological child to instrument for their parent's BMI.[17] The F-statistic from our first stage regression of mother's BMI on child's weight is 400, well in excess of the value 10 suggested by Bound et al. (1995); as such our analysis is unlikely to be affected by problems associated with weak instruments. The estimates from using child's weight as an instrument for self-reported mother's BMI are given in Table 6, where again the first column shows the true OLS estimate for comparison. In this case we see that the IV results predict that a 5 point increase in BMI reduces income by €7000, compared to a reduction of €4150 based on the true conditional mean.

In the comparisons so far we have assumed away the problem of endogeneity by imposing $E[\epsilon X^*] = 0$, in which case the slope of the conditional mean corresponds to the causal effect of BMI. However, in the event that this assumption is false the OLS estimate and the IV estimate in Table 7 are not strictly comparable, since the latter may adjust for endogeneity as well as measurement error. In other areas of labour economics repeated measures are often used as instruments when one is concerned about measurement error; for instance in the returns to education literature one may have access to the primary respondent's self-reported education, as well as a sibling's report on the respondent's education. While we are not aware of this approach being adopted in the research on obesity, we can nevertheless illustrate the consequences of using such instruments in the presence of measurement error by artificially creating appropriate instruments. To do this we take a random sample from a distribution with mean zero and a variance equal to the observed variance of the error in the data. This sample is drawn independently of both the true value of BMI and income. This error is then added to the recorded BMI available in the survey. We do this three times to create three new simulated values of BMI (BMI1-BMI3), all which can be thought of as

Table 7 Estimated coefficients on BMI in income equations using IV approach (GUI)

Dependent variable household income	OLS	IV (using child's weight as instrument)
Recorded BMI	−.821	
	(.091)	
IV BMI		−1.40
		(.398)
N	6637	6637

Income is measured in 000's; Standard Error in brackets.

measures of BMI subjected to classical measurement error. In line with the earlier discussion we can think of these as a spouse's or sibling's estimate of the respondent's BMI. The availability of more than one additional measure allows us to consider the consequences of measurement error for over-identification tests for instrument validity.

With these simulated variables we run three new regressions. Firstly, we regress income on BMI1 to illustrate the standard effects of classical measurement error. We then use two additional measures (BMI2 and BMI3) as instruments for BMI1; by construction these repeated measures are valid instruments in the presence of classical measurement error and should return the conditional mean. We then use these same instruments for our actual self-reported BMI measure and compare the findings. The results are given in Table 8. The first column repeats the results for the true specification for convenience. The results in the second column illustrate the textbook attenuation bias associated with classical measurement error. In our example the downward bias is of the order of 10%. The third column shows how the availability of "valid" instruments helps overcome the attenuation bias when the measurement error is classical. In addition a Hansen test of instrument validity fails to reject the over identifying restrictions that result from having two instruments. However, the results in the fourth column show how the use of these "valid" instruments do not result in consistent estimates when used to instrument the self-reported measure of BMI. The IV estimate *overestimates* the size of the true effect by 18%, which is somewhat larger than the bias in the raw OLS estimate. Furthermore the result of the Hansen test warns against relying on over identifying test to detect the problems with instruments in the presence of non-classical measurement error. This reflects the low power of this test when none of the instruments are valid.

4.2 BMI and household income (NHANES III)

To examine whether the results established above are specific to Irish data we consider the relationship between income and BMI using US data. The reported income data in

Table 8 IV using simulated instruments (GUI)

Dependent variable household income	OLS	OLS with classical measurement error	IV with classical measurement error (using BMI2 and BMI3 as instruments)	IV with non-classical measurement error (using BMI2 and BMI3 as instruments)
Recorded BMI	−.82			
	(.091)			
BMI1		−.74		
		(.086)		
BMI1 (IV using BM2 and BMI3)			−.85	
			(.089)	
self-reported BMI (IV using BMI2 and BMI3)				−.97
				(.101)
Hansen Over-identification test			p-value = .85	p-value = .92
N	6637	6637	6637	6637

Income is measured in 000's; Standard Error in brackets.

Table 9 Estimated coefficients on BMI in income regressions (NHANESIII)

Dependent variable household income	OLS	OLS	CE	IV (using BMI2 and BMI3 as instruments)
Recorded BMI	−.235			
	(.056)			
Self-Reported BMI		−.263		
		(.062)		
Corrected BMI using CE			−.249	
			(.059)	
Self-reported BMI (IV using artificial instruments)				−.265
				(.064)
Hansen's over identification tests				p-value = .29
N	2199	2199	2199	2199

Income is measured in 000's; Standard Error in brackets.

NHANES are only available on a bracketed basis. As a result we have estimated the income model using both interval regressions and OLS with the midpoint of the bracket as the dependent variable. Since the key findings were the same with either approach we only report the results using the midpoints. The results are provided in Table 9 and are consistent with the findings we reported using the Irish data. The first column shows a negative and statistically significant relationship between BMI and income when recorded BMI is used as the explanatory variable. The second column shows that using self-reported BMI overstates the estimated relationship, with a bias of the order of 12%. The CE approach remains biased due to differential measurement error. In contrast to the Irish results however the bias from the CE estimator in the NHANES data is smaller than that of the OLS estimator. This is because in the NHANES data − $cov(X^*, u) = 4.84$ and $var(u) = 2.91$, so that part b of Proposition 1 is now applicable. Finally the last column of Table 9 shows the results obtained using the IV estimator. We only consider the results using our artificially created instruments which do not adjust for endogeneity.[18] Again the results for the IV estimator overstate the true effect, with the bias from IV being somewhat larger than from OLS or the CE approach.

As before we check the robustness of these findings to the choice of additional controls specified in the model. Table 10 provides the parameter estimates from the income equation for the extended specification using the same set of additional controls that we used in the GUI analysis; namely age, education, health status and marital status, a control for whether the respondent smokes and a control for English language proficiency. In keeping with the findings from the GUI analysis the inclusion of these additional controls does not change the relative size of the bias from the CE approach; in the NHANES data the bias from the CE approach is of the order of 6% with or without the additional controls.

5. Conclusion and discussion

Obesity imposes very large costs on both governments and individuals. As a result there is a growing concern over measured levels of obesity throughout the world. However, studies examining the individual costs of obesity typically rely on self-reported data to measure BMI. The use of self-reported BMI gives rise to potential problems of

Table 10 Estimated coefficients on BMI in income regressions obtained using CE approach with additional controls (NHANES III)

Dependent variable household income	OLS	OLS	CE
Recorded BMI	−.197		
	(.050)		
Self-reported BMI		−.221	
		(.055)	
Corrected BMI			−.209
			(.052)
Years of education	2.24	2.24	2.24
	(.129)	(.129)	(.129)
Age	−.03	−.026	−.026
	(.024)	(.024)	(.024)
English language	0.28	0.473	0.473
	(2.57)	(2.57)	(2.57)
Married	11.39	11.69	11.69
	(.655)	(.663)	(.663)
Poor health	−7.20	−7.16	−7.16
	(2.22)	(2.22)	(2.22)
Smoker	−3.95	−3.94	−3.94
	(.702)	(.701)	(.701)
N	2191	2191	2191

Income is measured in 000's; Standard Error in brackets.

measurement error which could bias any estimated relationships. This paper uses both Irish and US data to explore the nature of measurement error in self-reported BMI and to examine the consequences of this error when estimating the relationship between BMI and income. The results are consistent across both data sets. We find that self-reported BMI is subject to substantial measurement error and importantly this error deviates from classical measurement error in two distinct ways. Firstly the error exhibits a pronounced negative correlation with the true measure of BMI; secondly self-reported BMI contains information about outcomes even after conditioning on true BMI. In our analysis we show that these departures from classical measurement error cause the traditional estimators to overstate the relationship between BMI and income. Furthermore we show that popular alternatives estimators that have been adopted to address problems of measurement error in BMI, such as the CE approach and the IV approach, continue to exhibit significant biases. The estimated biases are of the order of 6% to 20% depending on the procedure and data used.

Researchers interested in using self-reported data for their analysis will therefore need to think beyond the alternatives typically used. One possibility would be to consider the stochastic imputation approach developed for missing data, which has been extended to tackle measurement error problems (e.g. Brownstone and Valletta 1996; Freedman et al. 2008). One can view this approach as an extension of the mean imputation adopted by the CE estimator. There are many possible ways of carrying out stochastic imputation. A simple starting point is to assume that the distribution of $X^*|X \sim N(\alpha X, \sigma^2)$. Stochastic imputation would then proceed in a number of steps. In keeping with the CE approach

the first step would involve fitting a regression model to the validation data to obtain estimates of both $\hat{\alpha}$ and $\hat{\sigma}^2$. One can then simulate new parameters α_* and σ_*^2 from their joint posterior distribution under a non-informative prior (see Little and Rubin (2002) pg. 114). One can then obtain a set of simulated data by sampling X_{IMP}^* from $N\left(\alpha_* X_{nv}, \sigma_*^2\right)$ and a set of imputed estimates by regressing Y on X_{IMP}^*. Multiple imputations are needed to properly account for the uncertainty in the imputation process and the resulting set of imputed estimates can be combined using standard multiple imputation combining rules. Extensions to this simple approach would involve, for instance, considering the role of the outcome variable in the imputation process. While this approach typically requires specialist software, such software is now being included in some popular statistical packages, making the approach more accessible. In addition, since there is some evidence that the observed measurement error in self-reported BMI is heteroscedastic it may be possible to exploit this feature of the error process in the estimation procedure (Rigobon 2003; Guo and Little 2011). A final possibility would be to forsake point estimation and instead consider estimating bounds for the parameters. Such an approach would need to account for both the correlation between the measurement error and the true value of BMI as well as the differential nature of the measurement error. None of the bounds currently available allow for both these possibilities. The use of multiple imputation, the role of heteroscedasticity and the possibility of bounding parameter estimates in the presence of non-classical differential measurement error all provide interesting avenues for further research.

Endnotes

[1]Recently Burkhauser and Cawley (2008) compared multiple measures of fatness and found that many important patterns, such as who is classified as obese, group rates of obesity, and correlations of obesity with social science outcomes, are all sensitive to the measure of fatness and obesity used (see also Johansson et al. 2009; Wada and Tekin 2010; Parks et al. 2011). While these findings are interesting it is not the focus of our paper. The overwhelming majority of studies continue to use BMI to measure of obesity. For this reason a detailed empirical analysis of the biases arising from the use of self-reported BMI is of considerable value.

[2]For an overview of the IV approach see Angrist and Krueger (2001).

[3]This approach is in keeping with Bound et al. (1994). However, unlike that study we do not require that the measurement error in BMI be uncorrelated with the stochastic component of the income generating function.

[4]Measurement error is said to be non-differential when the conditional distribution of Y given X and X^* is the same as that of Y given X^*. In this case X is said to be a surrogate for X^*.

[5]Bound et al. (2001) page 3738, discuss an extension of this approach that uses internal validation data to correct for differential measurement error. We have verified that this approach works in our sample but do not focus on this adjustment here. We focus on adjustments that do not require internal validation data. These circumstances arise more often in practice and the consequences of measurement error are more serious in these situations.

[6]Cawley (2004) acknowledges this point and for this reason argues that it is important to correct for measurement error prior to using IV estimation (to control for endogeneity). However, most papers in the literature do not adopt this approach and

simply instrument the self-reported measure. Furthermore, as shown above, the CE approach does not overcome the problem of differential measurement error so that the problems with IV that we discuss can still apply to these "corrected" measures.

[7]We conducted all our analysis both allowing for and ignoring the complex survey design of both the GUI and NHANES. This made no difference to the key findings reported in the paper.

[8]See for example Morgan et al. (2008) and Shiely et al. (2010) for Ireland, Elgar and Stewart (2008) for Canada, Villanueva (2001) for the United States and Spencer et al. (2002) for the U.K.

[9]The negative correlation between measurement error and the true value of BMI is consistent with other findings (see Shiely et al. 2010; Elgar et al. 2005; Villanueva 2001; Spencer et al. 2002).

[10]We have also estimated the models in the paper restricting the sample to mothers who were working at the time of the survey. Although this reduced the sample sizes substantially it made very little difference to the findings of the paper.

[11]We have also estimated all our income models using a semi-log form with the log of income as the dependent variable. All the conclusions from the analysis reported in the paper are robust to this change.

[12]While we acknowledge the fact that an individual's perceptions of themselves may be important, this is not the focus of this paper.

[13]When carrying out this test it is important to control for dependence between the two estimators arising from the fact that the two regressions are estimated using the same sample.

[14]Some authors propose using non-linear specifications of BMI in the first stage regression of the CE approach. In addition a number of authors run separate regressions for height and weight when implementing this CE approach. Although the quantitative results change somewhat depending on which approached is used our findings that CE approach does not eliminate the bias from self-reported BMI is robust to the manner in which the CE approach is implemented.

[15]This extended specification is very similar to the model estimated by Brunello and D'Hombres (2007) when examining the impact of body weight on wages in Europe.

[16]The estimated coefficients on the new additional controls are identical in columns two and three. This result is not interesting and simply reflects the linear specification of the first stage regression in the CE approach.

[17]We abstract from the issue of whether child's weight is a valid instrument and simply observe, as noted in the main text, that the use of such instruments is very popular in this literature.

[18]Since not all households in the NHANES data contained children, using a child's weight as an instrument for adult BMI reduced the sample size by over two-thirds.

References

Angrist J, Krueger A (1992) The effect of Age at school entry on education attainment: an application of instrumental variables with moments from two samples. J Am Stat Assoc 87(418):328–336

Angrist J, Krueger A (2001) Instrumental variables and the search for identification: from supply and demand to natural experiments. J Econ Perspect 15(4):69–86

Bauhoff S (2011) Systematic self-report bias in health data: impact on estimating cross-sectional and treatment effects. Health Serv Outcomes Res Method 11(1):44–53. doi:10.1007/s10742-011-0069-3

Bound J, Brown C, Duncan G (1994) Evidence on the validity of cross-sectional and longitudinal labor market data. J Labor Econ 12:345–368

Bound J, Jaeger DA, Baker RM (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J Am Stat Assoc 90:443–450

Bound J, Brown C, Mathiowetz N (2001) Measurement error in survey data. In: Heckman JJ, Leamer EE (eds) , vol 5, Handbook of Econometrics North-Holland, Amsterdam

Brownstone D, Valletta R (1996) Modelling earnings measurement error: a multiple imputation approach. Rev Econ Stat 78(4):705–717

Brunello G, d'Hombres B (2007) Does body weight affect wages? evidence from Europe. Econ Human Biol 5:1–19

Burkhauser R, Cawley J (2008) Beyond BMI: the value of more accurate measures of fatness and obesity in social science research. J Health Econ 27:519–529

Carroll RJ, Ruppert D, Stefanski L (1994) Measurement error in nonlinear models. Chapman and Hall, London

Cawley J (2000) An instrumental variables approach to measuring the effect of body weight on employment disability. Health Serv Res 35:1159–1179

Cawley J, Markowitz S, Tauras J (2004) Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation. J Health Econ 23:293–311

Cawley J (2004) The impact of obesity on wages. J Hum Resour 39:451–474

Cawley J, Meyerhoefer C (2012) The medical care costs of obesity: an instrumental variables approach. J Health Econ 31:219–230

Conor Gorber S, Temblay M, Moher D, Gorver B (2007) A comparison of direct vs self-reported measures for assessing height, weight and body mass index: a systematic review. Obes Rev 8:307–326

Davey Smith G, Sterne J, Fraser A, Tynelius P, Lawlor D (2009) The association between BMI and mortality using offspring BMI as an indicator of own BMI: large intergenerational mortality study. Brit Med J 339. doi:10.1136/bmj.b5043

De Sousa S (2012) Does size matter? a propensity score approach to the effect of BMI on labour market outcomes. , Available via https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=pej2008&paper_id=64

Dor A, Ferguson C, Langwith C, Tan E (2010) A heavy burden: the individual costs of being overweight and obese in the United States. Available via http://sphhs.gwu.edu/departments/healthpolicy/dhp_publications/pub_uploads/dhpPublication_35308C47-5056-9D20-3DB157B39AC53093.pdf

Elgar FJ, Stewart JM (2008) Validity of self-report screening for overweight and obesity. Evidence from the Canadian Community Health Survey. Can J Public Health 99(5):423–427

Elgar FJ, Roberts C, Tudor-Smith C, Moore L (2005) Validity of self-reported height and weight and predictors of bias in adolescents. J Adolesc Health 37(5):371–375

Freedman L, Midthune D, Carroll R, Kipnis V (2008) A comparison of Regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. Stat Med 27:5195–5216

Fuller W (1987) Measurement Error Models. Wiley & Sons, New York

Gottschalk P, Huynh M (2010) Are earnings inequality and mobility overstated? the impact of nonclassical measurement error. Rev Econ Stat 92(2):302–315

Guo Y, Little R (2011) Regression analysis with covariates that have heteroscedastic measurement error. Stat Med 30:2278–2294. doi:10.1002/sim.4261

Hyslop D, Imbens G (2001) Bias from classical and other forms of measurement error. J Bus Econ Stat 19(4):475–481

IASCO (2010) Obesity: Understanding and Challenging the Global Epidemic, International Association for the Study of Obesity 2009–2010 Report

Inoue A, Solon G (2010) Two-sample instrumental variables estimators. Rev Econ Stat 92(3):557–561

Johansson E, Bockerman P, Kiiskinen U, Heliovaara M (2009) Obesity and labour market success in Finland: the difference between having a high BMI and being fat. Econ Human Biol 7:36–45

Kaestner R, Grossman M (2009) Effects of weight on Children's educational achievement. Econ Ed Rev 28:651–661

Kim B, Solon G (2005) Implications of mean-reverting measurement error for longitudinal studies of wages and employment. Rev Econ Stat 87:193–196

Kline B, Tobias J (2008) The wages of BMI: Bayesian analysis of a skewed treatment response model with nonparametric endogeneity. J Appl Econ 23:767–793

Kortt M, Leigh A (2010) Does size matter in Australia? Econ Rec 86(272):71–83

Lindeboom M, Lundborg P, van der Klaauw B (2010) Assessing the impact of obesity on labor market outcomes. Econ Human Biol 8:309–319

Little R, Rubin D (2002) Statistical analysis with missing data. John Wiley & Sons, New Jersey

Lyles R, Kupper L (1997) A detailed evaluation of adjustment methods for multiplicative measurement error in linear regression with applications in occupational epidemiology. Biometrics 53:1008–1025

Morgan K, McGee H, Watson D, Perry I, Barry M, Shelley E, Harrington J, Molcho M, Layte R, Tully N, van Lente E, Ward M, Lutomski J, Conroy R, Brugha R (2008) SLAN 2007: Survey of Lifestyle, Attitudes & Nutrition in Ireland: Main Report. Department of Health and Children, Dublin, Available via http://www.dohc.ie/publications/pdf/slan07_report.pdf?direct=1

O'Neill D, Sweetman O, Van de gaer D (2007) The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility. J Econ Inequal 5(2):159–178

Parks J, Smith A, Alston M (2011) Quantifying Obesity in Economic Research: How Misleading is the Body Mass Index. Available via http://agecon.ucdavis.edu/people/faculty/aaron-smith/docs/BMI%20Article%20DRAFT%20%203-30-2011.pdf

Plankey M, Stevens J, Flegal K, Rust P (1997) Prediction equations do Not eliminate systematic error in self-reported body mass index. Obes Res 5(4):308–314

Pischke J (1995) Measurement error and earnings dynamics: some estimates from the PSID validation study. J Bus Econ Stat 13(3):305–314

Rigobon R (2003) Identification through Heteroscedasticity. Rev Econ Stat 85(4):777–792

Shiely F, Perry I, Lutomski J, Harrington J, Kelleher C, McGee H, Hayes K (2010) Misclassification patterns of measured and self-report based body mass index categories - findings from three population surveys in Ireland. BMC Public Health 10:560. doi:10.1186/1471-2458-10-560

Spencer EA, Appleby PN, Davey GK, Key TJ (2002) Validity of self-reported height and weight in 4808 EPIC-oxford participants. Public Health Nutr 5(4):561–565

Stommel M, Schoenborn C (2009) Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the NHANES & NHIS 2001–2006. BMC Public Health 9:421. doi:10.1186/1471-2458-9-421

Villanueva E (2001) The validity of self-reported weight in US adults: a population based cross-sectional study. BMC Public Health 1:11. doi:10.1186/1471-2458-1-11

von Hinke S, Smith G, Lawlor D, Propper C, Windemeijer F (2012) The effect of Fat mass on educational attainment: examining the sensitivity to different identification strategies. Econ Human Biol 10:415–418

Wada R, Tekin E (2010) Body composition and wages. Econ Human Biol 8:242–254

WHO (2009) Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. World Health Organisation, WHO Press, Available via http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf