

Ljungvall, Åsa; Gerdtham, Ulf-G.; Lindblad, Ulf

**Working Paper**

## Misreporting and Misclassification: Implications for Socioeconomic Disparities in Body-mass Index and Obesity

Working Paper, No. 2012:19

**Provided in Cooperation with:**

Department of Economics, School of Economics and Management, Lund University

*Suggested Citation:* Ljungvall, Åsa; Gerdtham, Ulf-G.; Lindblad, Ulf (2012) : Misreporting and Misclassification: Implications for Socioeconomic Disparities in Body-mass Index and Obesity, Working Paper, No. 2012:19, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/260044>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Working Paper 2012:19

Department of Economics  
School of Economics and Management

# Misreporting and Misclassification: Implications for Socioeconomic Disparities in Body-mass Index and Obesity

Åsa Ljungvall  
Ulf-G Gerdtham  
Ulf Lindblad

July 2012



**LUND**  
UNIVERSITY

# ***Misreporting and misclassification: Implications for socioeconomic disparities in body-mass index and obesity***

Ljungvall, Å.\*<sup>a,b</sup>, Gerdtham, U-G.<sup>a,b,c</sup>, Lindblad, U.<sup>d</sup>

<sup>a</sup>Department of Economics, Lund University, Sweden

<sup>b</sup>Health Economics & Management, Institute of Economic Research, Lund University, Sweden

<sup>c</sup>Centre for Primary Health Care Research, Lund University, Sweden

<sup>d</sup>Department of Public Health and Community Medicine, University of Gothenburg, Sweden

## **Abstract**

Body-mass index (BMI), sometimes calculated from objectively measured and sometimes from self-reported weight and height, has become the standard proxy for obesity in social science research. This study deals with the potential problems related to, first, relying on self-reported weight and height to calculate BMI (*misreporting*), and, second, the concern that BMI is a deficient measure of body fat and elevated health risks (*misclassification*). Using a regional Swedish sample, we analyze whether socioeconomic disparities in BMI are biased because of misreporting, and whether socioeconomic disparities in the risk of obesity are sensitive to whether BMI or waist circumference is used to define obesity. Education and two income measures are used as socioeconomic indicators.

Among women, different educational groups misreport differently, leading to underestimation of the education disparity when using self-reported information. Among men, misreporting is un-related to socioeconomic status, but misclassification is related to education. As a consequence, when estimating the risk of obesity defined using waist circumference, an educational gradient, which is not present when classifying men using BMI, arises. Taken together, female disparities appear more sensitive to whether weight and height are self-reported, whereas male disparities are more sensitive to definition of obesity.

**Keywords:** body-mass index, waist circumference, obesity, socioeconomic disparity, misreporting, misclassification

**JEL-codes:** I12, I14

---

\* Corresponding author. E-mail address: asa.ljungvall@nek.lu.se

## 1. INTRODUCTION

Obesity is nowadays recognized as an important public health concern, and considerable obesity-related research is being produced in different fields. A common feature in most of the obesity research in social sciences is that, despite its shortcomings, body-mass index (BMI, calculated as weight in kilos divided by height in meters squared,  $\text{kg/m}^2$ ) has become the standard proxy for body fat and is the most widely used indicator for obesity (Burkhauser and Cawley, 2008; Kuczmarski, 2007). Many times BMI is the only body measure available, sometimes calculated from self-reported, and sometimes from objectively measured, weight and height. BMI has the important advantage of being relatively easy and cheap to collect, especially if weight and height are self-reported. However, it is well-known that the self-reported weight and height are misreported in a way that tends to understate BMI and obesity prevalence.

In statistical terms, misreporting is an example of measurement error, which in a regression may introduce bias in the estimated parameters, where the direction and severity of the bias depends on the model specification and how the measurement error is related to all other variables in the model (Bound, Brown, and Mathiowetz, 2001). The measurement error problem can be overcome by using an external dataset with more accurate data to quantify the error and thereafter correct for it in the primary dataset (Bound *et al.*, 2001). Studying the relationship between wages and obesity, Cawley (2004) uses such a strategy to correct self-reported weight and height in a U.S. dataset. The relationship between the self-reported and measured information in the validation data is used to adjust the self-reported data in the primary dataset, and these adjusted values are used in the analysis, instead of the original self-reported values. Many other studies that use U.S. datasets with only self-reported height and weight follow this method (Cawley, Moran, and Simon, 2010; Chou, Grossman, and Saffer, 2004; Fletcher *et al.*, 2010; Lakdawalla and Philipson, 2009; Ruhm, 2005), and Gil and Mora (2011) and Mora and Gil (2012) apply the same method on Spanish data.

The current study does not aim at correcting measurement error in any primary dataset. Instead it focuses explicitly on the misreporting behavior and misclassification per se, with particular attention to socioeconomic factors. It illustrates whether and how the shortcomings of the common use of self-reported weight and height to calculate BMI, and the common use of BMI to define obesity, matter for socioeconomic disparities. The analysis is based on regional Swedish data and consists of two parts. The first part deals with *misreporting* in BMI, with the specific purpose to analyze whether misreporting behavior varies systematically across socioeconomic groups. If it does, socioeconomic gradients based on self-reported data will be biased. We find that among women there are significant differences in reporting behavior across education, and women with post-secondary education underreport BMI to a smaller extent than lower educated, leading to underestimation of socioeconomic disparities in BMI when using self-reported information. Among men, we find no evidence of systematic differences across socioeconomic groups.

Previous studies give some evidence that misreporting may vary systematically across socioeconomic groups. Some of these studies control for measured weight and height, or BMI, whereas others do not. Consequently, some studies analyze whether different socioeconomic groups report differently *given the same true level* of the body measure, whereas others analyze whether different groups report differently overall – a difference that is formalized and discussed in more detail in Section 3. Nyholm et al. (2007) use a Swedish regional dataset (partly the same as is used in the current study) and report that there is a slight tendency for men in the middle, and women in the highest, educational group to report more accurate values of weight, height, and BMI calculated from these. Dekkers et al. (2008) use a sample of overweight employees in the Netherlands. Controlling for measured height and weight, misreporting is smaller in the higher educational group. On the other hand, controlling for a broad set of covariates, including measured BMI, Gil and Mora (2011) find no systematic differences across education and individual deprivation in the misreporting of weight or height in a Spanish

dataset. None of these studies discuss the implications of their results in terms of socioeconomic disparities or other biases. Finally, using a Swedish regional dataset collected in 1984-1985, Boström and Diderichsen (1997) find some evidence for differences in misreporting by occupation. They conclude that this misreporting results in underestimated socioeconomic disparities in the risk of obesity among women, and in overestimated disparities in overweight and obesity among men.

The second part of the analysis in this study deals with *misclassification* and goes beyond using BMI to define obesity. The standard is to classify an individual as obese if  $BMI \geq 30$ . In a U.S. context, Burkhauser and Cawley (2008) and Burkhauser, Cawley, and Schmeiser (2009) note that obesity prevalence is much higher when using a more direct measure of body fat, instead of BMI, to define obesity. Further, when defining obesity using the more direct measure, the negative correlation between employment and obesity increases for men, but not for women (Burkhauser and Cawley, 2008).

In the current study we use waist circumference as an indicator of *abdominal*, or *central*, obesity, and as an alternative measure of elevated health risk. Central obesity is considered to provide an independent prediction of risk beyond BMI, in particular among individuals with  $BMI < 35$ , and is therefore a valuable complement to BMI (Kuczmarski, 2007). Differently to BMI, waist circumference takes fat distribution into account. High-risk central obesity is defined as a waist circumference of more than 88 cm for women, and more than 102 cm for men (Andersson and Fransson, 2011; Kuczmarski, 2007). We use these cut-off points, and explore whether *misclassification*, defined as being classified as obese according to the waist circumference definition but not according to the commonly used BMI definition, is systematically related to socioeconomic status. Although *misclassification* points towards one definition being straightforwardly more appropriate than the other, this is not necessarily so in the current case. BMI and waist circumference are both proxies, and they are proxies for different underlying

measures, total body fat and excess abdominal body fat, respectively (Kuczmarski, 2007). Nevertheless we use the word misclassification. The logic is that being centrally obese, as measured by waist circumference, but still having a relatively low BMI, means that the elevated risk associated with this situation is overlooked when focusing on BMI. Using misclassification in this way, we find that misclassification is related to education for men, and unrelated to socioeconomic status for women.

As a summarizing step, we finally estimate socioeconomic gradients in obesity for three different definitions of obesity: waist circumference and  $\text{BMI} \geq 30$  calculated from self-reported and measured weight and height, respectively. As expected from our previous results, the largest difference in the estimated socioeconomic gradient appears when moving from defining obesity based on self-reports to measured weight and height for women, whereas the largest difference for men appears when moving from obesity defined using BMI to obesity defined using waist circumference.

Taken together, this study contributes by shedding light on misreporting and misclassification patterns. Whether there exists systematic misreporting and misclassification across socioeconomic groups is of interest for a wide range of obesity research where self-reported weight and height are used, and where obesity is defined as  $\text{BMI} \geq 30$ . In the current study, we pay particular attention to systematic differences across socioeconomic groups. As many datasets contain only self-reported height and weight as body measures, and because it is important to track and explore disparities, self-reported values and BMI are used as the best available option, and socioeconomic disparities in BMI and/or obesity are analyzed based on these values (Baum and Ruhm, 2009; Costa-Font and Gil, 2008; de Saint Pol, 2009; García Villar and Quintana-Domeque, 2009; Heineck, 2006; Ljungvall and Gerdtham, 2010; van der Pol, 2011; Reinhold and Jürges, 2010). The current study explains and shows how the systematic

misreporting behavior affects disparities in BMI, and how the misclassification affects disparities in the risk of obesity.

## 2. METHODS

### 2.1 Misreporting

Analyzing socioeconomic disparities, the relationship of interest is whether and how *actual* BMI differs across socioeconomic groups, as specified in the following linear regression framework:

$$BMI\_meas_i = \alpha^{meas} + age_i * \beta^{meas} + x_i * \gamma^{meas} + e_i \quad (1)$$

$BMI\_meas$  is BMI calculated from objectively measured height and weight for individual  $i$ .  $age_i$  is a row vector that consists of 46 dummy variables, one for each age between 31 and 76, keeping individuals who are 30 years old as reference.  $x_i$  is a row vector of socioeconomic variables and  $e_i$  is a residual term.  $\gamma^{meas}$  are the parameters of main interest and reveal whether BMI differs across socioeconomic groups. We define  $\gamma^{meas}$  as the “true” socioeconomic gradient, with the logic that it is true in the sense that it is estimated from true BMI.

When true BMI is not available, self-reported data are used instead:

$$BMI\_self_i = \alpha^{self} + age_i * \beta^{self} + x_i * \gamma^{self} + \varepsilon_i \quad (2)$$

where  $BMI\_self$  refers to BMI calculated from self-reported weight and height, and all other notation is as before. Despite estimating Equation 2, the parameters of interest are  $\gamma^{meas}$ . Hence, it is relevant to ask whether  $\gamma^{meas} = \gamma^{self}$ , and thereby whether  $\gamma^{self}$  are unbiased estimates of the “true” disparities. The following equation is used to test whether  $\gamma^{meas} = \gamma^{self}$ :



$$BMI\_self_i - BMI\_meas_i = \alpha_3 + \mathbf{age}_i * \boldsymbol{\beta}^{total} + \mathbf{x}_i * \boldsymbol{\gamma}^{total} + \tau_i \quad (3)$$

where notation is as before, and  $BMI\_self - BMI\_meas$  is defined as misreporting. If  $\boldsymbol{\gamma}^{total} \neq 0$ ,  $\boldsymbol{\gamma}^{meas}$  and  $\boldsymbol{\gamma}^{self}$  are significantly different and hence, there is bias in the estimated disparities based on the self-reported data.

The potential bias in  $\boldsymbol{\gamma}^{self}$  consists of a direct and an indirect effect of socioeconomic status. To see this, it is useful to express misreporting as a function of true BMI:

$$BMI\_self_i - BMI\_meas_i = \alpha_4 + \rho BMI\_meas_i + \mathbf{age}_i * \boldsymbol{\beta}^{direct} + \mathbf{x}_i * \boldsymbol{\gamma}^{direct} + r_i \quad (4)$$

where notation is as before.  $\rho$  reveals whether misreporting is related to the level of true BMI.  $BMI\_self - BMI\_meas < 0$  means that BMI calculated from self-reported weight and height is underreported, and  $\rho < 0$  implies that underreporting increases with the true level of BMI. If  $\boldsymbol{\gamma}^{direct} > 0$ , underreporting decreases with socioeconomic status, *given the same level of true BMI and age*. This is referred to as the direct effect of socioeconomic status on misreporting. To see also the indirect effect, substitute Equation 1 into the right hand side of Equation 4:

$$\begin{aligned} BMI_{self_i} - BMI_{meas_i} &= \\ &= (\rho \alpha^{meas} + \alpha_4) + \mathbf{age}_i * (\boldsymbol{\beta}^{meas} \rho + \boldsymbol{\beta}^{direct}) + \mathbf{x}_i * (\boldsymbol{\gamma}^{meas} \rho + \boldsymbol{\gamma}^{direct}) + (\rho e_i + r_i) \end{aligned} \quad (5)$$

where notation is as before. Equation 5 shows that the total misreporting attributable to the socioeconomic status variable  $\mathbf{x}_k$  can be decomposed into  $(\boldsymbol{\gamma}_k^{meas} \rho + \boldsymbol{\gamma}_k^{direct})$ . Hence, the total difference related to socioeconomic status consists of the direct effect shown in Equation 4,  $\boldsymbol{\gamma}_k^{direct}$ , and an indirect effect  $\boldsymbol{\gamma}_k^{meas} \rho$ . The indirect effect is a combination of the “true” gradient,  $\boldsymbol{\gamma}_k^{meas}$ , and the effect of measured BMI on misreporting behavior. Hence, the indirect effect

appears if true BMI varies systematically with socioeconomic status, and if misreporting additionally is related to true BMI.

Previous studies that analyze misreporting in self-reported weight and height do not discuss and distinguish between the direct and indirect effect. Some of them measure the total effect, as in Equation 3, and some of them measure the direct effect through an approach similar to Equation 4. In this study, we consider both the total difference across socioeconomic groups, and the decomposition into indirect and direct effects. We estimate Equation 1 and 2 to compare the resulting disparities when using objectively measured and self-reported data, respectively. We then estimate Equation 3 and test the null hypothesis that  $\gamma^{total} = 0$ . Following Equation 5, the potential bias in  $\gamma^{self}$  can be decomposed into a direct and an indirect effect. The direct effect,  $\gamma^{direct}$ , is estimated in Equation 4. The indirect effect is estimated in Equation 4 ( $\rho$ ) and 1 ( $\gamma^{meas}$ ). The equations are estimated by OLS with heteroskedasticity robust standard errors, assuming that the residual terms are normally distributed with a zero mean, and are estimated for men and women separately.<sup>1</sup>

## 2.2 Misclassification

The second part of the analysis deals with misclassification, defined as having a waist circumference above the cut-off point for high risk of adverse health outcomes (88 cm for women, 102 cm for men), but not being categorized as obese based on objectively measured BMI, where obesity is defined as BMI $\geq$ 30 for both men and women.

The relationship between misclassification and socioeconomic status is estimated by OLS in a linear probability model (LPM), with heteroskedasticity robust standard errors:

---

<sup>1</sup> Equations 1-5 could also be specified with the BMI variables transformed to their log values in order to capture a non-linear relationship between BMI and socioeconomic status as well as between true BMI and the absolute size of misreporting, for example if the absolute size of underreporting increases at an increasing rate with the true level of BMI. The results from this alternative analysis are similar to the results in the main analysis, and are available from the authors upon request.

$$\Pr(misclassified_i) = \alpha_6 + \mathbf{age}_i * \boldsymbol{\beta}^{miss} + \mathbf{x}_i * \boldsymbol{\delta}^{miss} + \varepsilon_i \quad (6)$$

where notation is as before.  $\boldsymbol{\delta}$  are the parameters of main interest and indicates whether socioeconomic status is related to the probability to be misclassified.

Finally, to see directly whether different definitions of obesity results in different socioeconomic gradients, we estimate the risk of being obese as a function of age and socioeconomic status for three different definitions of obesity:

$$\Pr(obese_i) = \alpha_7 + \mathbf{age}_i * \boldsymbol{\beta}^{obese} + \mathbf{x}_i * \boldsymbol{\delta}^{obese} + \varepsilon_i \quad (7)$$

where *obese* is defined using BMI $\geq$ 30 based on self-reported or measured weight and height, or using waist circumference. Also Equation 7 is estimated by OLS with robust standard errors.<sup>2</sup>

### 3. DATA AND VARIABLES

Up to date, in Sweden there is no nationally representative dataset that contains measured information about weight and height. The current analysis therefore uses a regional sample, collected between 2001 and 2005 in a region in the South part of Sweden. The dataset consists of two surveys. One was conducted between 2001 and 2004 in the municipality of Vara (participation rate 81 percent), and the other in 2004 and 2005 in the nearby municipality of Skövde (participation rate 70 percent). In each survey, individuals in the age between 30 and 76 years were randomly selected from the population in strata by age and sex, and invited to two visits at a health care center. During the first visit, participants answered a questionnaire including questions about their height and weight. When they came back for the second visit,

---

<sup>2</sup> We also estimated Equation 6 and 7 by logit and probit models, calculating both average marginal effects and marginal effects at the mean, giving very similar results irrespective of estimation method, and irrespective of how marginal effects are calculated. These results are available from the authors upon request.

their height and weight were measured, which the participants did not know when they filled out the questionnaire during the first visit. Also waist circumference was measured at this time.

To the survey data we link register data on education and income from Statistics Sweden. Individuals are classified into four educational groups: up to eleven years of schooling (*educ1*), two or three years of high school (*educ2*), up to three years of university or other post-secondary education (*educ3*), and at least three years of post-secondary education (*educ4*). 37 observations miss education register information. 30 of these have self-reported information which is used instead.

The income measure is household disposable income per consumption unit.<sup>3</sup> This variable is provided for the year when the individual participated in the survey, and is used as current income. In addition, the income variable is provided also for 1985 and 1995. Using the average of disposable income in 1985, 1995, and current income we construct an alternative, and a more stable, long-term, income measure.<sup>4</sup> All incomes are adjusted for inflation by the consumer price index and are measured in 2005 year's prices.

Pregnant women are excluded, and after losing observations due to missing information about income (n=31 in total), education (n=6), and self-reported (n=136) or measured (n=2) height and/or weight, the final sample consist of 1329 men and 1302 women.

All individuals were supposed to do their second visit at the health care center 14 days after the first visit. However, Table I shows that there is indeed some variation in the number of days

---

<sup>3</sup> Different household members have different consumption weights depending on age and household size and composition. The consumption weights are as follows: The first adult in the household has weight 1.16, the second co-habiting adult has weight 0.76, other persons above 18 years old have weight 0.96, children 11-17 years old have weight 0.76, children 4-10 years old have weight 0.66, and children 0-3 years old have weight 0.56.

<sup>4</sup> For individuals with missing information on income in 1985 (n=53) or 1995 (n=1), the average between current income and income in 1995 or 1985 is used. Individuals with missing information on income in both 1985 and 1995 (n=24) are excluded from the analysis.

between the visits. If dates for the first and second visits are recorded correctly, the number of days between the visits generally varies between 0 and 60 days, with some additional outlying observations. The median is indeed the intended 14 days for both men and women.

[Table I about here]

As the time period between the visits increases, the risk that the observed difference between BMI calculated from self-reported and measured information is an actual weight difference, and not a misreport, increases (height reasonably do not change in the age groups included in the analysis). If actual weight increases are driving the recorded misreporting, we would expect misreporting to increase with the number of days between the first and second visit. For re-visits up to 63 days after the first visit, Figure 1 shows no such apparent increase in the misreporting as the number of days increases. Further, Table II reports that regressing misreporting on the number of days between the first and second visit, and its square, gives no significant results, which lends further support to the idea that the observed differences between self-reported and measured weight and height are misreports, and not driven by actual increases in weight.

[Figure I about here]

[Table II about here]

## **4. RESULTS**

### **4.1 Descriptive statistics**

Table III reports final sample descriptive statistics. As expected, BMI calculated from measured height and weight is higher than BMI based on self-reports. In line with previous literature, underreporting is related to the level of BMI. Stratifying by BMI classification (underweight:  $BMI < 18.5$ , normal weight:  $18.5 \leq BMI < 25$ , overweight:  $25 \leq BMI < 30$ , and obese:  $BMI \geq 30$ ) shows

that underweight women (there are no underweight men in the sample) on average overreport BMI. Both men and women in the three other BMI statuses on average underreport. Obese individuals underreport more than overweight, who underreport more than normal weight people.

Defining obesity as  $BMI \geq 30$ , obesity prevalence increases by four percentage points for both men and women when using measured values instead of self-reports. Notably, when defining obesity using waist circumference instead, prevalence increases to 35 percent among women. Among men, the increase is less pronounced.

[Table III about here]

Regarding misclassification, Figure 2 illustrates the relationship between BMI (calculated from objectively measured height and weight) and waist circumference. The lower right square of each graph represent the misclassified observations. 15 percent of female observations, and six percent of male, are misclassified according to the definition we use. Fewer both men (three percent) and women (two percent) are allocated in the upper left corners, and thus have a relatively high BMI but a slim waistline, indicating that the person is muscular. Hence, even though the difference is less pronounced among men, for both sexes, the risk of misclassification as defined in this study is larger than the risk of wrongly categorizing muscular individuals as obese when using BMI to evaluate the obesity status.

[Figure II about here]

## 4.2 Misreporting

Table IV and V report the results from the misreporting analysis. The tables report results from five different models. Model I includes three indicator variables for level of education in the  $x_i$

vector. Model II includes the log of current income, and Model III includes the log of long-term income. Model IV includes both education and current income, and Model V combines education and long-term income.

The first column reports the results from Equation 1, where measured weight and height are used to calculate BMI to estimate “true” gradients. The second column reports the socioeconomic disparities in case only self-reported information would be available (Equation 2). The third column reports the results from estimation of Equation 3 and gives evidence to whether the estimates based on the self-reported data in column 2 are biased. The fourth column reports the results from estimation of Equation 4 and answers whether there is any direct effect of socioeconomic status and age on the total bias.

For women (Table IV), estimation of Model I shows, as expected, that there are educational disparities when using BMI calculated from both measured and self-reported weight and height. Based on measured information (column 1), BMI among women in the two highest educational groups are about 1.6 and 1.8 index points lower, respectively, compared to a woman in the lowest educated group. These estimates are statistically significantly larger (i.e. more negative) than the ones estimated from self-reported data. Hence, the estimated gradient based on self-reported values is biased towards zero. According to column 4, there is a statistically significant direct effect of education on the bias for the highest education group: given the same level of true BMI, women in this group report weight and height in a way that results in less underreporting of BMI compared to the lowest educated. Because women in the higher educational groups also tend to have lower BMI than the lower educated, and because women with lower BMI underreport BMI to a lesser extent than women with higher BMI, there is also an indirect effect of education on the bias. Following Equation 5, the indirect effect is a combination of  $\rho$  and  $\gamma^{meas}$ , which for the highest educational group is  $(-0.059)*(-1.761)=0.104$ , corresponding to about 27 percent of the total bias. For the second highest educational group

(*educ3*), the indirect effect is relatively more important and corresponds to about 36 percent of the total bias. Adding income in Model IV and V reduces the education effects slightly, but they are still significantly larger using BMI calculated from measured instead of self-reported weight and height.

In Model II, using BMI calculated from measured weight and height results in a negative income gradient which is significant at the 10 percent level, but once controlling for education, it loses significance (Model IV). Using self-reported information instead gives a somewhat smaller, and insignificant, effect, but the difference between the two estimates is not significant (column 3). Model III uses the more long-term income as indicator of socioeconomic status. Here again there are income disparities both when using measured and self-reported weight and height to calculate BMI, and the effect is larger than for current income in Model II. According to column 1, a 10 percent increase in income is related to a 0.15 index points lower BMI. As for education, the effect is larger (i.e. more negative) when using measured information, and this difference is statistically significant (column 3). The direct effect of income accounts for 67 percent of the total bias associated with this variable (column 5), and is significant at the 10 percent level (column 4). Adding education in Model V, the long-term income effect reduces in size and significance, and the bias in the gradient based on self-reports disappears. Overall, education seems to be stronger associated to BMI than income in this sample.

[Table IV about here]

The results for men are reported in Table V. According to Model I, there is an education gradient also among men, where higher educated men have lower BMI, but the educational differences are smaller than among women. Compared to a man with less than eleven years of schooling (*educ1*), a man in the highest educational group has about one index point lower BMI. The income gradients in Model II and III are negative, but small and insignificant. Controlling for



both income and education (Model IV and V) gives very similar education disparities as Model I. Hence, as for women, education is the strongest obesity-related socioeconomic variables in this sample, whereas income is less important.

However, differently to the female results, only one estimated gradient is significantly different between the specification with self-reported and measured information (column 3). In Model V the “true” long-term income effect is smaller than in the self-reported case, and the difference is significant at the ten percent level. Hence, men with higher long-term income underreport to a lesser extent than men with lower income. Model III, not controlling for education, shows the same thing but without significance. However, the relationship between long-term income and BMI is weak in both Model III and V, and despite the significant difference in reporting behavior across long-term income, the income gradients are insignificant. Nevertheless, had the long-term income and BMI relationship been stronger, using self-reported weight and height to calculate BMI, would result in an overestimation of the long-term income effect.

Further, in Model I, IV and V, there is a tendency towards a *negative* direct effect for the highest educational group; given the same level of true BMI, men in the highest education group tend to underreport BMI to *larger* extent than the lowest educated. In Model V this effect is significant at the ten percent level. In general, however, male socioeconomic disparities in BMI calculated from self-reported weight and height seem to be less biased than corresponding female disparities.

[Table V about here]

#### **4.3 Misclassification**

Table VI reports the results from estimation of Equation 6. Among women (columns 1-5) there does not seem to be any systematic patterns of misclassification across education or income.

Compared to the lowest educated, the probability to be misclassified is 3.4-4.8 percentage points larger for women in the highest education group, but the difference is insignificant. The difference between the three lowest educational groups is very small.

In contrast, among men (columns 6-10), there is a rather large education effect. In Model I, men in the two highest educational groups are four and 5.6 percentage points less likely to be misclassified than the lowest educated, respectively. The difference is statistically significant at the ten percent level for the second highest educational group, and at the five percent level for the highest educated group. The size of the effect is rather unaffected by adding income to the model (Model IV and V). The income parameter estimates are negative, but small and insignificant, in all models.

[Table VI about here]

The finding of systematic variation across education in the probability of misclassification implies that the education gradient in obesity is underestimated when defining obesity based on BMI compared to a definition based on waist circumference. Table VII illustrates this implication by reporting the results from estimating Equation 7, where obesity is defined in three different ways. Columns 4-6 show that when moving from obesity defined based on BMI to defining obesity based on waist circumference, a negative education effect for the two highest educational groups evolves also for men. The size of this gradient is 6-9 percentage points. The effects are significant at the five to ten percent level, but are yet not statistically significantly different from the effect estimated when defining obesity based on true BMI.

For women, the most noticeable change in the education disparity occurs when moving from BMI calculated from self-reports to BMI calculated from measured information as definition of obesity (column 1 and 2). When moving further, to obesity defined using waist circumference,

the *educ3* estimate remains similar in size, and women in this group are about 10 percentage points less likely to be obese compared to the lowest educated. The *educ4* estimate reduces in size, from 13-14 percentage points to 6-8 percentage points. In Model IV and V, it also loses significance, but yet the differences to the estimates in column 2 are statistically significant only in Model IV. The income effects are somewhat larger for the waist circumference definition, although these differences are not statistically significant.

[Table VII about here]

## 5. CONCLUSIONS

This study deals with the potential problems related to, first, relying on self-reported weight and height to calculate BMI (*misreporting*), and, second, the concern that BMI is a deficient measure of body fat and elevated health risks (*misclassification*). We analyze how these potential problems affect estimates of socioeconomic disparities in BMI and obesity. We find that women with higher education misreport less than lower educated women. Accordingly, when analyzing socioeconomic disparities in BMI derived from self-reported weight and height, the resulting (particularly educational) disparities are underestimated and biased towards zero, compared to disparities derived from observational data. Among men, we find no evidence of reporting heterogeneity across income or education.

In the misclassification part, we use waist circumference as an alternative measure of health risk. Descriptive statistics show that female obesity prevalence increases considerably when applying this alternative definition. Male obesity prevalence also increases, but less markedly. Hence, these raw statistics reveal that focus on BMI as definition of obesity, may understate the actual risks and problems, especially among women. We also find that lower educated men tend to be misclassified to a larger extent than higher educated men. As a consequence, when estimating the risk of obesity defined using waist circumference, an educational gradient, which

is not present when classifying men using BMI, arises. Among women, misclassification does not appear to be systematically related to socioeconomic status.

Taken together, in short, the conclusion is that socioeconomic disparities among women are more sensitive to whether weight and height are self-reported, whereas male disparities are more sensitive to whether BMI or waist circumference are used to define obesity.

As for validation data studies, the generalizability of the results from this study is limited by the regional characteristic of the data set. For the results from the misreporting analysis in this study to be useful in a broader context, for example in order to draw conclusions at the national level, the distribution of BMI calculated from measured height and weight, conditional on the distribution of the corresponding self-reported values, age, and socioeconomic status, must be the same in the both populations. Likewise, for the misclassification results to be valuable in a broader context, the relationships between central obesity, BMI, age and socioeconomic status must be similar in both contexts. It is difficult to judge whether these assumptions are likely to hold. In the region where the data used in this study were collected, the fraction of individuals with at least three years of post-secondary education is somewhat smaller, and the fraction with very low education is somewhat higher, compared to the average of Sweden (Statistics Sweden, 2005). According to a report comparing health outcomes across Swedish municipalities, based on survey data collected between 2006 and 2008, obesity prevalence (defined as  $BMI \geq 30$  calculated from self-reports) among 18-80 years old men and women in the region under consideration in this study, is somewhat higher than in the country as an average, although confidence intervals are indeed overlapping (SALAR, 2009). However, these factors do not necessarily imply that misreporting behavior is any different than in the rest of Sweden. Overall, despite the regional character and the generalization limitation, we believe that the results from this study contribute with valuable and interesting insights about the nature and consequences of misreporting in BMI and misclassification of obesity.

## References

- Andersson D, Fransson A. 2011. *Kalorier kostar: en ESO-rapport om vikten av vikt*. ESO-rapport (Vol. 2011:3). Finansdepartementet, Regeringskansliet: Stockholm. *In Swedish*.
- Baum C L, Ruhm C J. 2009. Age, socioeconomic status and obesity growth. *Journal of Health Economics* **28** : 635–648.
- Boström G, Diderichsen F. 1997. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *International Journal of Epidemiology* **26** : 860–866.
- Bound J, Brown C, Mathiowetz N. 2001. Measurement error in survey data. In J. Heckman, E. Leamer (Eds.), *Handbook of Econometrics* (1<sup>st</sup> ed., volume 5, pp. 3705-3843). Springer-Verlag: New York.
- Burkhauser R V, Cawley J. 2008. Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics* **27** : 519–529.
- Burkhauser R V, Cawley J, Schmeiser M. D. 2009. The timing of the rise in U.S. obesity varies with measure of fatness. *Economics & Human Biology* **7** : 307–318.
- Cawley J. 2004. The impact of obesity on wages. *Journal of Human Resources* **39** : 451–474.
- Cawley J, Moran J, Simon K. 2010. The impact of income on the weight of elderly Americans. *Health Economics* **19** : 979–993.
- Chou S Y, Grossman M, Saffer H. 2004. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics* **23** : 565–587.
- Costa-Font J, Gil J. 2008. What lies behind socio-economic inequalities in obesity in Spain: A decomposition approach. *Food Policy* **33** : 61–73.
- de Saint Pol T. 2009. Evolution of obesity by social status in France, 1981–2003. *Economics & Human Biology* **7** : 398–404.
- Dekkers J C, van Wier M F, Hendriksen I J M, Twisk J W R, van Mechelen W. 2008. Accuracy of self-reported body weight, height and waist circumference in a Dutch overweight working population. *BMC Medical Research Methodology* **8**.
- Fletcher J M, Frisvold D, Tefft N. 2010. Can soft drink taxes reduce population weight? *Contemporary Economic Policy* **28** : 23–35.

- García Villar J, Quintana-Domeque C. 2009. Income and body mass index in Europe. *Economics & Human Biology* **7** : 73–83.
- Gil J, Mora T. 2011. The determinants of misreporting weight and height: The role of social norms. *Economics & Human Biology* **9** : 78–91.
- Heineck G. 2006. Height and weight in Germany, evidence from the German Socio-Economic Panel, 2002. *Economics & Human Biology* **4** : 359–382.
- Kuczmarski R J. 2007. What is obesity? Definitions matter. In S. Kumanyika, R. Brownson (Eds.), *Handbook of obesity prevention: A resource for health professionals* (pp. 25–44). Springer Science + Business Media, LLC: New York.
- Lakdawalla D, Philipson T. 2009. The growth of obesity and technological change. *Economics & Human Biology* **7** : 283–293.
- Ljungvall Å, Gerdtham U-G. 2010. More equal but heavier: A longitudinal analysis of income-related obesity inequalities in an adult Swedish cohort. *Social Science & Medicine* **70** : 221–231.
- Mora T, Gil J. 2012. Peer effects in adolescent BMI: Evidence from Spain. *Health Economics* doi: 10.1002/hec.2817.
- Nyholm M, Gullberg B, Merlo J, Lundqvist-Persson C, Råstam L, Lindblad U. 2007. The validity of obesity based on self-reported weight and height: Implications for population studies. *Obesity* **15** : 197–208.
- Reinhold S, Jürges H. 2010. Secondary school fees and the causal effect of schooling on health behavior. *Health Economics* **19** : 994–1001.
- Ruhm C J. 2005. Healthy living in hard times. *Journal of Health Economics* **24** : 341–363.
- SALAR - Swedish Association of Local Authorities and Regions. 2009. *Öppna jämförelser 2009 – Folkhälsa*. In collaboration with the Swedish National Institute of Public Health and the National Board of Health and Welfare. SALAR: Stockholm. *In Swedish*.
- Statistics Sweden. 2005. *Statistics Sweden's Statistical Database*:  
<http://www.ssd.scb.se/databaser/makro/start.asp> [10 April 2012]. Tables: *Population 16-74 years of age by municipality, level of education, and sex* and *Population 16-74 years of age by level of education and sex*. Data for 2005.
- van der Pol M. 2011. Health, education and time preference. *Health Economics* **20** : 917–929.

## TABLES

**Table I.** Descriptive statistics for the number of days between the first and second visit at the health care center.

	<b>WOMEN</b>	<b>MEN</b>
	n=1213	n=1213
min	0 (n=4)	0 (n=7)
mean	21.78	19.77
max	751	393
Std. Dev.	40.55	22.68
<b>Percentiles</b>		
5th	2	2
25th	8	7
50th	14	14
75th	23	24
95th	62	59

Note: The number of observations differs from the final sample used in the main analysis because of missing information on the date for the first visit. Observations are included in the main final sample regardless of this information and the number of days between the visits.

**Table II.** Misreporting as a function of the number of days between the first and second visit at the health care center. OLS regressions.

<b>WOMEN</b>						
	(1)	(2)	(3)	(4)	(5)	(6)
days	-0.001 (0.001)	-0.000 (0.002)	-0.001 (0.002)	-0.002 (0.002)	-0.008 (0.007)	-0.004 (0.007)
days^2		-0.000 (0.000)	0.000 (0.000)		0.000 (0.000)	0.000 (0.000)
days and days^2 joint significance (p-value)		0.566	0.296		0.493	0.230
Controls for age, education, and income	no	no	yes	no	no	yes
Max no of days	751	751	751	63	63	63
No. of observations	1213	1213	1213	1158	1158	1158
<b>MEN</b>						
days	0.000 (0.001)	-0.001 (0.002)	-0.002 (0.002)	-0.003 (0.002)	-0.013* (0.008)	-0.011 (0.007)
days^2		0.000 (0.000)	0.000 (0.000)		0.000 (0.000)	0.000 (0.000)
days and days^2 joint significance (p-value)		0.102	0.229		0.202	0.149
Controls for age, education, and income	no	no	yes	no	no	yes
Max no of days	393	393	393	63	63	63
No. of observations	1213	1213	1213	1165	1165	1165

Notes: Heteroskedasticity robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.  
Dependent variable: BMI\_self-BMI\_meas.



**Table III.** Descriptive statistics. Final sample.

	<b>Women</b>					<b>Men</b>				
	<i>n</i>	<i>Mean</i>	<i>st. dev</i>	<i>min</i>	<i>max</i>	<i>n</i>	<i>Mean</i>	<i>st. dev</i>	<i>min</i>	<i>max</i>
<b>BMI</b>										
bmi_meas	1302	26.59	5.24	15.78	52.60	1329	26.88	3.62	18.53	49.17
bmi_self	1302	25.86	5.00	15.43	51.90	1329	26.31	3.44	14.88	48.42
bmi_self - bmi_meas	1302	-0.73	1.07	-10.93	4.06	1329	-0.57	1.11	-12.98	3.19
bmi_self - bmi_meas if bmi_meas < 18.5	11	0.44	0.65	-0.35	1.80					
bmi_self - bmi_meas if $18.5 \leq \text{bmi\_meas} < 25$	593	-0.42	0.71	-3.38	2.13	417	-0.19	0.82	-3.27	3.19
bmi_self - bmi_meas if $25 \leq \text{bmi\_meas} < 30$	408	-0.85	1.12	-10.93	4.06	680	-0.58	1.03	-11.26	2.81
bmi_self - bmi_meas if bmi_meas $\geq 30$	290	-1.22	1.36	-7.27	1.51	232	-1.22	1.43	-12.98	1.31
<b>weight (kg)</b>										
weight_meas	1302	72.89	15.02	38.90	152.00	1329	86.30	12.73	51.80	152.30
weight_self	1302	71.18	14.52	39.00	150.00	1329	84.71	12.35	41.00	150.00
weight_self - weight_meas if low weight	435	-1.03	1.56	-5.80	5.90	446	-0.93	2.75	-31.90	8.90
weight_self - weight_meas if middle weight	435	-1.49	2.17	-15.30	12.20	441	-1.48	2.97	-34.90	10.40
weight_self - weight_meas if high weight	432	-2.61	3.04	-17.70	3.70	442	-2.35	3.01	-24.50	9.20
<b>height (centimeter)</b>										
height_meas	1302	165.57	6.16	146.00	185.00	1329	179.14	6.75	156.00	199.00
height_self	1302	165.86	5.99	148.00	198.00	1329	179.37	6.59	157.00	199.00
height_self - height_meas if short	471	0.67	2.45	-10.00	40.00	445	0.59	2.07	-9.00	13.00
height_self - height_meas if middle length	432	0.32	1.38	-3.00	10.00	489	0.17	1.64	-10.00	7.00
height_self - height_meas if tall	399	-0.19	1.14	-5.00	4.00	395	-0.10	1.41	-8.00	9.00
<b>obesity (%)</b>										
obese_meas	1302	0.22	0.42	0.00	1.00	1329	0.17	0.38	0.00	1.00
obese_self	1302	0.18	0.39	0.00	1.00	1329	0.13	0.34	0.00	1.00
<b>waist circumference</b>										
wasit circumference (centimeter)	1302	85.07	13.35	56.00	164.00	1329	94.73	9.94	63.00	145.00
central obesity (%)	1302	0.35	0.48	0.00	1.00	1329	0.20	0.40	0.00	1.00
<b>Age and socioeconomics</b>										
age	1302	46.95	11.40	30.00	76.00	1329	47.12	11.60	30.00	76.00
educ1	1302	0.27	0.45	0.00	1.00	1329	0.28	0.45	0.00	1.00
educ2	1302	0.47	0.50	0.00	1.00	1329	0.52	0.50	0.00	1.00
educ3	1302	0.15	0.36	0.00	1.00	1329	0.12	0.33	0.00	1.00
educ4	1302	0.11	0.31	0.00	1.00	1329	0.07	0.26	0.00	1.00
income (SEK)	1302	129359	49142	18811	565532	1329	138509	53972	1	548130
average income (SEK)	1302	103466	30897	26625	375739	1329	112673	32700	31958	316323

**Table IV.** Misreporting analysis. Women.

Dependent variable:	BMI (measured information)	BMI (self-reported information)	BMI (self-reported) - BMI (measured)	BMI (self-reported) - BMI (measured)	Direct effect (%)
	Equation 1 (1)	Equation 2 (2)	Equation 3 (3)	Equation 4 (4)	(5)
<b>MODEL I</b>					
bmi_meas $\rho$				-0.059*** (0.007)	
educ2	0.019 (0.377)	0.162 (0.367)	0.143 (0.094)	0.144 (0.091)	101
educ3	-1.627*** (0.466)	-1.358*** (0.450)	0.269** (0.115)	0.172 (0.109)	64
educ4	-1.761*** (0.466)	-1.367*** (0.450)	0.394*** (0.107)	0.290*** (0.101)	73
R-squared	0.105	0.090	0.090	0.165	
<b>MODEL II</b>					
bmi_meas $\rho$				-0.061*** (0.007)	
ln_income	-0.695* (0.391)	-0.597 (0.382)	0.098 (0.081)	0.056 (0.080)	57
R-squared	0.088	0.076	0.080	0.160	
<b>MODEL III</b>					
bmi_meas $\rho$				-0.060*** (0.007)	
ln_inc_av	-1.522*** (0.538)	-1.246** (0.518)	0.276** (0.115)	0.185* (0.111)	67
R-squared	0.092	0.078	0.083	0.162	
<b>MODEL IV</b>					
bmi_meas $\rho$				-0.059*** (0.007)	
educ2	0.056 (0.385)	0.198 (0.375)	0.142 (0.094)	0.145 (0.092)	102
educ3	-1.578*** (0.472)	-1.310*** (0.456)	0.267** (0.115)	0.174 (0.110)	65
educ4	-1.666*** (0.483)	-1.275*** (0.468)	0.391*** (0.111)	0.292*** (0.106)	75
ln_income	-0.283 (0.402)	-0.274 (0.394)	0.009 (0.085)	-0.008 (0.083)	-91
R-squared	0.105	0.090	0.090	0.165	
<b>MODEL V</b>					
bmi_meas $\rho$				-0.059*** (0.007)	
educ2	0.139 (0.390)	0.263 (0.380)	0.123 (0.097)	0.132 (0.095)	107
educ3	-1.451*** (0.480)	-1.211*** (0.464)	0.240** (0.118)	0.154 (0.113)	64
educ4	-1.501*** (0.488)	-1.150** (0.471)	0.351*** (0.116)	0.263** (0.110)	75
ln_inc_av	-0.988* (0.561)	-0.826 (0.542)	0.162 (0.123)	0.104 (0.119)	64
R-squared	0.107	0.092	0.091	0.166	

Notes: Heteroskedasticity robust standard errors in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Number of observations: 1302. A constant and age effects (46 age dummies) are included in all models, but not shown in table.

**Table V.** Misreporting analysis. Men.

Dependent variable:	BMI (measured information)	BMI (self-reported information)	BMI (self-reported) - BMI (measured)	BMI (self-reported) - BMI (measured)	Direct effect (%)
	Equation 1 (1)	Equation 2 (2)	Equation 3 (3)	Equation 4 (4)	(5)
<b>MODEL I</b>					
bmi_meas $\rho$				-0.093*** (0.010)	
educ2	-0.379 (0.255)	-0.330 (0.241)	0.049 (0.085)	0.014 (0.080)	28
educ3	-0.781** (0.327)	-0.661** (0.315)	0.120 (0.107)	0.048 (0.102)	40
educ4	-1.006*** (0.385)	-1.043*** (0.365)	-0.037 (0.106)	-0.130 (0.099)	352
R-squared	0.062	0.055	0.055	0.141	
<b>MODEL II</b>					
bmi_meas $\rho$				-0.092*** (0.010)	
ln_income	-0.147 (0.167)	-0.106 (0.158)	0.041 (0.036)	0.027 (0.034)	67
R-squared	0.056	0.049	0.054	0.140	
<b>MODEL III</b>					
bmi_meas $\rho$				-0.092*** (0.010)	
ln_inc_av	-0.083 (0.382)	0.137 (0.351)	0.221 (0.136)	0.213* (0.127)	97
R-squared	0.056	0.048	0.056	0.142	
<b>MODEL IV</b>					
bmi_meas $\rho$				-0.093*** (0.010)	
educ2	-0.378 (0.255)	-0.330 (0.242)	0.049 (0.085)	0.014 (0.080)	28
educ3	-0.767** (0.327)	-0.652** (0.315)	0.115 (0.106)	0.044 (0.101)	38
educ4	-0.984** (0.387)	-1.030*** (0.367)	-0.045 (0.105)	-0.137 (0.098)	301
ln_income	-0.104 (0.169)	-0.063 (0.162)	0.041 (0.035)	0.031 (0.034)	76
R-squared	0.062	0.055	0.055	0.141	
<b>MODEL V</b>					
bmi_meas $\rho$				-0.093*** (0.010)	
educ2	-0.393 (0.257)	-0.363 (0.244)	0.029 (0.083)	-0.007 (0.078)	-25
educ3	-0.806** (0.331)	-0.721** (0.319)	0.085 (0.102)	0.010 (0.097)	12
educ4	-1.037*** (0.389)	-1.118*** (0.368)	-0.081 (0.101)	-0.177* (0.095)	219
ln_inc_av	0.155 (0.390)	0.374 (0.359)	0.219* (0.132)	0.233* (0.122)	107
R-squared	0.062	0.055	0.058	0.144	

Notes: Heteroskedasticity robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Number of observations: 1329. A constant and age effects (46 age dummies) are included in all models, but not shown in table.

**Table VI.** Misclassification analysis. Women and men. Linear probability models.

Dependent variable: misclassification										
<b>Women (n=1302)</b>						<b>Men (n=1329)</b>				
mean of dependent variable: 0.149						mean of dependent variable: 0.058				
	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V	(6) Model I	(7) Model II	(8) Model III	(9) Model IV	(10) Model V
educ2	-0.005 (0.028)			0.001 (0.028)	0.001 (0.028)	-0.023 (0.019)			-0.023 (0.019)	-0.023 (0.019)
educ3	-0.003 (0.036)			0.004 (0.036)	0.005 (0.037)	-0.040* (0.021)			-0.038* (0.021)	-0.040* (0.022)
educ4	0.034 (0.040)			0.048 (0.042)	0.046 (0.042)	-0.056** (0.024)			-0.053** (0.024)	-0.055** (0.025)
ln_income		-0.032 (0.028)		-0.041 (0.030)			-0.017 (0.014)		-0.015 (0.014)	
ln_inc_av			-0.033 (0.037)		-0.045 (0.040)			-0.017 (0.023)		-0.004 (0.025)
R-squared	0.056	0.055	0.055	0.057	0.057	0.052	0.049	0.048	0.053	0.052

Notes: Heteroskedasticity robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. A constant and age effects (46 age dummies) are included in all models, but not shown in table.

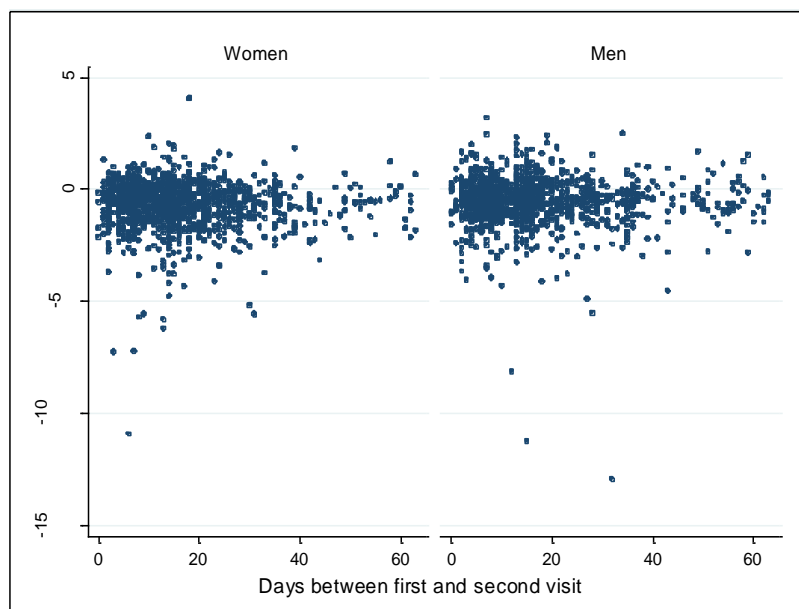
**Table VII.** Estimation of socioeconomic disparities in the risk of obesity for different definitions of obesity. Women and men. Linear probability models.

Dependent variable:	Women (n=1302)			Men (n=1329)		
	BMI≥30 self-reported	BMI≥30 measured	waist circumference > 88 cm	BMI≥30 self-reported	BMI≥30 measured	waist circumference > 102 cm
Mean:	0.184	0.223	0.352	0.131	0.175	0.200
	(1)	(2)	(3)	(4)	(5)	(6)
<b>MODEL I</b>						
educ2	0.032	0.010	0.003	-0.011	-0.018	-0.026
	(0.030)	(0.032)	(0.036)	(0.025)	(0.028)	(0.030)
educ3	-0.071**	-0.105***	-0.106**	-0.032	-0.037	-0.067*
	(0.033)	(0.036)	(0.044)	(0.032)	(0.038)	(0.038)
educ4	-0.095***	-0.143*** (a)	-0.087*	-0.045	-0.058	-0.092**
	(0.034)	(0.036)	(0.048)	(0.038)	(0.041)	(0.043)
<b>MODEL II</b>						
ln_income	-0.058*	-0.052*	-0.087**	-0.026	-0.012	-0.029
	(0.030)	(0.031)	(0.038)	(0.019)	(0.022)	(0.021)
<b>MODEL III</b>						
ln_inc_av	-0.088**	-0.083*	-0.120**	-0.040	-0.003	-0.022
	(0.041)	(0.044)	(0.051)	(0.036)	(0.042)	(0.044)
<b>MODEL IV</b>						
educ2	0.037	0.013	0.012	-0.011	-0.018	-0.026
	(0.030)	(0.032)	(0.036)	(0.024)	(0.028)	(0.030)
educ3	-0.065*	-0.102***	-0.094**	-0.029	-0.036	-0.064*
	(0.034)	(0.037)	(0.045)	(0.033)	(0.038)	(0.038)
educ4	-0.082**	-0.137*** (a)	-0.064 (b)	-0.040	-0.056	-0.087**
	(0.036)	(0.038)	(0.050)	(0.038)	(0.042)	(0.043)
ln_income	-0.038	-0.020	-0.069*	-0.024	-0.010	-0.025
	(0.031)	(0.033)	(0.039)	(0.020)	(0.023)	(0.022)
<b>MODEL V</b>						
educ2	0.040	0.015	0.015	-0.008	-0.019	-0.026
	(0.030)	(0.032)	(0.036)	(0.025)	(0.028)	(0.030)
educ3	-0.060*	-0.098***	-0.089**	-0.027	-0.039	-0.066*
	(0.034)	(0.038)	(0.046)	(0.033)	(0.038)	(0.039)
educ4	-0.078**	-0.133*** (a)	-0.062	-0.038	-0.060	-0.092**
	(0.036)	(0.038)	(0.050)	(0.038)	(0.042)	(0.044)
ln_inc_av	-0.062	-0.039	-0.094*	-0.032	0.010	-0.002
	(0.043)	(0.046)	(0.053)	(0.037)	(0.043)	(0.046)

Notes: Heteroskedasticity robust standard errors in parentheses.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1. A constant and age effects (46 age dummies) are included in all tables, but not shown in table. (a): the estimate is statistically significantly (p<0.05) different from the corresponding estimate where BMI≥30 calculated from self-reported weight and height are used to define obesity (column 1 for women and column 4 for men). (b): the estimate is statistically significantly (p<0.1) different from corresponding estimate where BMI≥30 calculated from measured weight and height is used to define obesity (column 2 for women and column 5 for men).

## FIGURES

**Figure I.** Scatterplots of the number of days between the first and second visit at the health care center and the difference in BMI calculated from self-reported and measured weight and height. For observations with fewer than 65 days between visits.



**Figure II.** Relationship between BMI (calculated from measured height and weight) and waist circumference. Women and men.

