

Burke, Mary A.; Carman, Katherine G.

Working Paper

You can be too thin (but not too tall): Social desirability bias in self-reports of weight and height

Working Papers, No. 16-15

Provided in Cooperation with:

Federal Reserve Bank of Boston

Suggested Citation: Burke, Mary A.; Carman, Katherine G. (2016) : You can be too thin (but not too tall): Social desirability bias in self-reports of weight and height, Working Papers, No. 16-15, Federal Reserve Bank of Boston, Boston, MA

This Version is available at:

<https://hdl.handle.net/10419/171769>

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.



You Can Be Too Thin (But Not Too Tall): Social Desirability Bias in Self-Reports of Weight and Height

Mary A. Burke and Katherine G. Carman

Abstract

Previous studies of survey data for the United States and other countries find that on average women tend to understate their body weight, while on average both men and women overstate their height. Social norms have been posited as a potential explanation for misreporting of weight and height, but researchers disagree on the validity of that explanation. This paper is the first to present a theoretical model of self-reporting behavior for weight and height that explicitly incorporates social desirability bias. The model generates testable implications that can be contrasted with predictions based on alternative explanations for self-reporting errors. Using data from the National Health and Nutrition Examination Survey (NHANES) from 1990–2010, we find that self-reporting patterns for both weight and body mass index (BMI) offer robust evidence of social desirability bias, such that reports are biased (from both sides) towards social norms. The BMI norm inferred for women lies squarely within the range considered “healthy” by public health officials, while the BMI norm inferred for men lies just above this healthy range. Lack of awareness of one’s current body weight may explain the presence of large (negative) self-reporting errors among those with very high values of examined weight, but the evidence of social desirability bias is robust to this alternative explanation over most of the weight distribution. Social desirability bias in self-reporting of height is observed primarily among those of below-average height and no clear height norms are discernible. The framework also helps to explain previous findings that the degree of self-reporting bias in weight depends on the survey mode.

JEL Classifications: D03, I12, I18

Mary A. Burke is a senior economist in the research department at the Federal Reserve Bank of Boston. Her e-mail address is mary.burke@bos.frb.org. Katherine G. Carman is an economist at the RAND Corporation and a professor at the Pardee RAND Graduate School. Her e-mail address is kcarman@rand.org.

This paper presents preliminary analysis and results intended to stimulate discussion and critical comment. The views expressed herein are those of the authors and do not indicate concurrence by the Federal Reserve Bank of Boston, or by the principals of the Board of Governors, or the Federal Reserve System.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at <http://www.bostonfed.org/economic/wp/index.htm>.

This version: August 15, 2016

1. Introduction

It is widely recognized that self-reported values of body weight and height are often inaccurate.¹ In the United States as well as in a number of other countries, body mass index (BMI) values based on self-reported weight and height tend to be too low on average, and population obesity rates may be significantly lower when these are calculated using self-reported data rather than by using direct measurements of weight and height (Yun et al. 2006). Despite the importance for public health policy of obtaining accurate estimates of obesity, many local and/or national governments continue to rely on self-reported data because the cost of collecting such data is much lower than the cost of collecting objective data on weight and height. Self-reported values are used in the United States to estimate state-level and county-level obesity rates, and are used in many countries (including Italy and France) in order to estimate national obesity rates.

Given the ongoing reliance on self-reported weight and height data, a significant literature has arisen that analyzes and seeks to correct for biases that arise when using such data in economic and/or epidemiological analysis. A separate and smaller strand of the literature on self-reported weight and height seeks to identify the reasons for misreporting. A leading explanation is that self-reports are biased in the direction of prevailing social norms for weight and height. This is an example of “social desirability bias,” which has been observed in numerous other contexts—such as in studies that find that voting rates based on surveys are significantly overstated, that people either overstate or understate their incomes in surveys, and that individuals are apt to underreport engaging in risky behaviors such as drug use.²

To date, the evidence concerning social desirability bias in self-reporting of weight and height is suggestive but inconclusive. The debate over the reasons for misreporting of weight and height suffers from the lack of a precise conceptual framework describing how social norms or other factors might influence self-reporting behavior. This paper addresses this gap by presenting a simple model in which self-reporting of weight (and height) involves a tradeoff between the desire to report a socially desirable value and the desire to give a truthful report. The incentive to provide an accurate measure stems from the presence of intrinsic and extrinsic costs of dishonesty, where the latter depend on the extent to which the self-reported value can be validated by others.

Our model generates testable predictions that are strongly supported in the NHANES data from 1999–2010 that contains both self-reported and independently examined values of weight and height, plus

¹For studies of the United States, see Connor Gorber et al. 2007, Cawley et al. 2015, Engstrom et al. 2003, Cawley 2004.

²Evidence of social desirability in self-reports of voting is found in Holbrook and Krosnick (2010) and Karp and Brockington (2005). Evidence pertaining to income is described in Tourangeau and Yan (2007) and Kelly (2015). Evidence on biased self-reporting of drug use is given in Krumpal (2013) and van de Mortel (2008).

numerous controls. In particular, individuals whose true weight exceeds a given weight threshold or norm tend to understate their weight, those below the norm tend to overstate their weight, and the absolute value of self-reporting errors increases in the distance between true weight and the social norm. Self-reported BMI values derived from self-reported weight and height exhibit the same qualitative patterns as self-reported weight, even though self-reported height offers more limited evidence of social desirability bias. The data suggest that the extent of social desirability bias in self-reporting of weight and BMI is mild to moderate rather than severe, indicating that individuals also care about giving an accurate report. Within the model's framework, the concern for accuracy may be heightened by aspects of the NHANES survey design that increase the psychic costs of providing biased reports.

We also discuss alternative explanations for self-reporting errors and their empirical implications. Among the alternatives we consider, the explanation that involves an individual who is unaware of his/her current body weight is the only one that might generate stylized patterns similar to those derived from our model of social desirability bias. Indeed, we find that the presence of large negative reporting errors among those with very high examined weights might reflect such a lack of awareness, especially among women. On a related note, among men weighing over 400 pounds, underreported body weight values may reflect weight gains occurring between the time when weight is self-reported and the time when weight is examined. However, even after controlling for multiple proxies for weight awareness and weight changes over time, the evidence of/for social desirability bias appears robust over most of the weight distribution.

The empirical analysis allows us to infer social norms for BMI. We find that these values are broadly consistent with public health standards in that these values lie within or close to the "normal" BMI range specified by the U.S. Centers for Disease Control (CDC).³ However, the inferred BMI norm for women (22 kg/m^2) is lower than the inferred BMI norm for men (26 kg/m^2). This latter finding agrees with prior sociological evidence describing ideal or socially normative physiques that are gender-specific: a slender ideal prevails among women, while a larger, more muscular build is seen as ideal for men.⁴

Our findings have implications for the design of surveys that collect self-reported values of weight and height. To the extent that biases reflect the subjects' ignorance of their own body weight, participants might be provided with a scale and encouraged to weigh themselves prior to the interview, although this might result in selective nonparticipation. However, our findings suggest that this measure alone would not eliminate systematic misreporting tendencies. Previous evidence (for example, Pinkston 2015) indicates that self-reports of weight and height collected in person tend to be more accurate than those collected over the phone. In our framework, this difference is explained by the fact that the psychic cost to

³For clinical guidelines see National Institutes of Health (1998).

⁴The literature on ideal physiques is too voluminous to cite comprehensively. Examples pertaining to ideals for men include Leit et al. 2001 and Leit et al. 2002; examples pertaining to women include Bordo 2003 and Groesz et al. 2002.

an individual of providing an inaccurate report increases based on the extent to which the report can be validated by the persons administering the survey. On the other hand, the survey mode (phone or in-person) would not be expected to affect the accuracy of self-reports if inaccuracy merely reflected ignorance or inattention.

The presence of social desirability bias suggests that surveys might seek ways to minimize this bias. One way could be to heighten the salience of the costs related to dishonesty, which in our theoretical framework work against social desirability bias. In addition to providing some degree of validation or accountability of self-reported data, survey instruments should (and often do) prompt individuals to consider the benefits to researchers of providing honest reports. To the extent that ignorance of one's true weight contributes to reporting errors, individuals might be asked to weigh themselves in private on a home scale, and be supplied with a scale if they don't have one, before providing a self-report of body weight.

Section 2 discusses the related literature. Section 3 presents the theoretical model and describes the empirical strategy. Section 4 describes the data. Section 5 presents the main results. Section 6 discusses the robustness of the main results to alternative explanations, and Section 7 concludes.

2. Related Literature

A large literature has examined the accuracy of self-reported values of weight, height, and body mass index, using data such as the National Health and Nutrition Examination Survey (NHANES), which includes both direct measures and self-reported measures. This literature is reviewed in Connor Gorber et al. (2007).⁵ Reviewing over 60 articles, they find that individuals on average overreport their height and underreport their weight, and as a result tend to underreport their BMI. However, Connor Gorber et al. find that the degree of misreporting varies significantly across individuals with factors including age, race, sex, and objective weight status. In particular, women are more likely to underestimate their weight than men, and men are more likely to overreport their height.⁶

Many previous studies have speculated or asserted that self-reported values of body weight and height are influenced by social norms for weight and height, and a smaller set of papers (discussed below) have attempted to test this hypothesis in various ways. Such an influence constitutes an example of social desirability bias, a phenomenon by which, in an attempt to make a good impression on the surveyor, survey responses are skewed towards what the survey-taker perceives is a socially desirable answer.

⁵A large related literature seeks to construct correction equations that adjust for biases associated with self-reporting. See, for example, Connor Gorber et al. (2008), Cawley (2004), Stommel and Schoenborn (2009), and Courtemanche et al. (2015).

⁶Among individual papers, see, for example, Engstrom et al. (2003), Roland (1990), and Cawley (2002).

Using data from Spain, Gil and Mora (2011) find that social norms of weight—defined as the average weight of a reference group—affect individuals’ misreporting of their own weight. In particular, individuals are less likely to underreport their weight when their reference group has a higher average weight: the hypothesis being that higher weight among peers results in a higher weight norm and therefore less underreporting. However, it must be noted that these findings derive from a single region of Spain and may not hold either throughout Spain or in the United States.

Stommel and Osier (2013) find that, between NHANES III (1988-1994) and NHANES 2005-2008, the tendency to underreport BMI declined significantly among obese individuals. They argue that this change in self-reporting behavior reflects a shift in social attitudes between the two surveys that reflect the increase over the same period in average measured BMI in the U.S. population.

Using a subset of the same unrestricted NHANES data that we employ, covering the 2003-2010 period, Cawley et al. (2015) identify roughly the same stylized pattern in self-reporting errors for weight (and also for height and BMI) that we describe. They note that weight-reporting errors are non-classical, but they do not take a strong stand either way on whether social norms contribute to the non-classical pattern. Instead, their paper seeks to identify the direction of bias when estimating the associations between body weight and health outcomes using self-reported weight and height data.

Courtemanche et al. (2015) provides a method of correcting for self-reporting errors in weight and height that are robust to differences in self-reporting errors across surveys, provided that the surveys refer to the same underlying population. The method is important because, for example, the extent of underreporting of weight appears to be greater in self-reported data from the Behavioral Risk Factor Surveillance System (BRFSS) than in self-reported data from NHANES. This difference in bias is thought to reflect the fact that self-reports of weight and height are collected over the phone in the BRFSS and in person in the NHANES. Similarly, Pinkston (2015) finds in NLSY data that a given individual tends to report a lower weight when interviewed over the phone than in person.

Ambwani and Chmielewski (2013) use a social desirability score measured in an independent psychological assessment and relate this score to misreporting behavior for body weight. Among college women in the United States, they find that those who underreport body weight have a higher average social desirability score than those who overreport their weight. Among college men, average social desirability scores are not significantly different between overreporters and underreporters. The authors thus infer that social desirability influences weight reporting among women but not among men. We will show, however, that this difference-in-means test is inappropriate because social desirability bias may result in either overreporting or underreporting of weight. Furthermore, the samples used in the study are small ($N < 200$) and are not broadly representative of the U.S. population.

DeAndrea et al. (2012) conduct three experiments involving self-reporting of body weight and height. The first experiment finds that heightening the salience of social desirability—by prompting an individual to consider her most attractive physical features or behaviors—leads to more severe distortions of BMI among women but not among men. The second experiment showed that having subjects weigh themselves prior to reporting the values dramatically increased the accuracy of reports, and the third showed that introducing accountability—informing subjects that they would be weighed after self-reporting—also reduced distortions. DeAndrea et al. conclude that social desirability bias is a potential factor in explaining the misreporting of weight and height, but that some lack of awareness of one’s own body weight might be required for biases to occur among some individuals. While this study also provides suggestive evidence that social desirability influences self-reporting behavior for weight and height, like Ambwani and Chmielewski (2013) it was conducted using a small sample (N=150) of college students.

Using survey data from Ireland, Brestoff, Perry, and Van den Broeck (2011) question the hypothesis that the desire to conform to social norms helps to explain reporting biases for weight and height. The authors posit that if social norms mattered, individuals who classify themselves as overweight (based on a direct, qualitative survey question) would be more likely to understate their true weight than individuals who classify their weight as “about right,” since the former reveal a greater awareness of social norms regarding weight. After controlling for measured weight, Brestoff, Perry, and Van den Broeck find, on the contrary, that those individuals who classify themselves as overweight are less likely to underreport their weight than those who classify themselves as having a normal weight. The authors claim that their findings challenge the relevance of social norms to self-reporting behavior. They speculate, based on indirect evidence, that a lack of awareness of one’s own current body weight, due to a combination of infrequent weighing and weight gain over time, might constitute a better explanation for underreporting by overweight individuals.

All of the preceding work related to social desirability bias in self-reported values of weight and height relies solely on empirical analysis, and different authors have tested for the influence of such bias in very disparate ways, making it difficult to weigh the different types of evidence against each other. The debate over the influence of social norms on self-reporting of weight and height has been clouded by the lack of a conceptual framework within which to assess the empirical findings in a consistent way. In contrast, our model offers a precise description of the influence of norms on self-reporting behavior and generates testable implications that can be contrasted with predictions based on alternative mechanisms for self-reporting errors. Furthermore, our model generates predictions concerning the entire distribution of reporting errors, not just average behaviors, enabling richer tests of the theory than tests based on means alone. We also examine the potential role of infrequent weighing in explaining self-reporting errors and

find that evidence of social desirability bias is robust to this alternative hypothesis over most of the weight distribution, especially among men.

3. Conceptual Framework

3.1 Model of Self-Reporting Behavior with Social Desirability Bias

The model describes an individual's self-reported body weight value as a function of an individual's true weight and an exogenously given social norm for body weight. The model can be adapted to describe self-reporting behavior for height, as discussed below. In the models pertaining to weight, we abstract from an individual's height for simplicity. However, we control for height differences in the empirical analysis of weight self-reporting.

The basic idea of the model is that, when reporting one's body weight in a survey, an individual faces a tradeoff between reporting truthfully and reporting a value that agrees with the social norm. Both the individual's true weight and the social norm are taken as given, and the norms are treated as universal within each sex. This tradeoff arises because an individual gets a psychic payoff that is decreasing in both the distance of the reported weight value from the social norm and, separately, in the distance of the reported value from the true value, which we assume is known to the individual. Reporting a value close to the norm (but which departs from the true value) is an example of social desirability bias which, as noted above, has been observed in previous surveys in reference to a variety of behaviors. The motivation to be truthful may derive from an internalized norm of honesty, or from the fact that reporting an obviously false value could lead to embarrassment. This latter possibility is particularly relevant in the case of the NHANES because in that survey self-reported values of weight and height are elicited during an in-person interview. In such a setting the interview subject may perceive that the interviewer would be able to judge the accuracy of self-reported values of weight and height based on visual inspection.

The utility derived from an individual's self-reported body weight is assumed to be linearly separable from other sources of utility, and is parameterized as follows:

$$U(w^{SR}; w^T, M) = -\alpha(w^{SR} - w^T)^2 - \beta(w^{SR} - M)^2. \quad (1)$$

In the above, w^{SR} denotes self-reported weight, which is the choice variable; w^T denotes the individual's true weight, which is assumed to be a fixed value that is observed accurately by the individual, and M denotes the social norm for body weight. The norm is treated as given, but is allowed to differ between women and men. We assume that the parameters α and β are fixed, strictly positive numbers; we do not assume anything about their relative size. The expression $-\alpha(w^{SR} - w^T)^2$ indicates that, all else constant, utility decreases quadratically (and symmetrically) in the difference between self-reported

weight and true weight. Similarly, the expression $-\beta(w^{SR} - M)^2$ implies that utility decreases quadratically in the difference between self-reported weight and the social norm. If we were to let $\beta=0$, utility would be maximized by reporting one's true weight, and if we were to let $\alpha=0$, utility would be maximized by reporting a weight equal to the norm. When both α and β are strictly positive, the optimal value of self-reported weight, denoted w^{SR*} , is given as follows:

$$w^{SR*} = \frac{\alpha w^T + \beta M}{\alpha + \beta}. \quad (2)$$

The expression above shows that the optimal self-reported weight is simply a weighted average of one's true weight and the social norm. If α is greater than β , w^{SR*} will be closer to the true weight than to the norm, and vice-versa if β exceeds α . Subtracting the true weight, w^T , from both sides of equation (2), we obtain an expression for the “optimal” self-reporting error—defined as the (signed) difference, $w^{SR*} - w^T$, as follows:

$$w^{SR*} - w^T = \frac{\beta(M - w^T)}{\alpha + \beta}. \quad (3)$$

Equation (3) shows that the optimal error is proportional to the difference between the group weight norm and the individual's true weight. This means that if one's true weight exceeds the norm, the optimal self-reported weight (w^{SR*}) will be strictly less than the true weight (w^T), and the self-reporting error will have a negative sign—the individual is said to “underreport” his/her true weight. If the true weight is less than the norm, however, the optimal self-reported weight will strictly exceed the true weight and the individual “overreports.” The farther above (below) the norm is true weight, the greater the absolute extent of underreporting (overreporting).⁷

Taking the social norm as given and universal, equation (3) implies that the optimal self-reporting error is a linear function of an individual's true weight, where the coefficient on true weight equals $(-\beta/(\alpha + \beta))$. Since both α and β are strictly positive, this value lies strictly between 0 and -1 . Holding α fixed, larger values for β imply a steeper negative slope—that is, the greater the relative utility benefit of reporting a weight close to the social norm (rather than a truthful weight), the steeper the (negative) relationship between true weight and self-reporting error will be. Figure 1 shows the predicted self-reporting errors as a function of true weight for a scenario involving a relatively weak preference for reporting a value close to the norm (a small value of β) and for a separate scenario involving a stronger norm preference (a larger value of β). The weight norm is arbitrarily set at 132 pounds in both cases.

⁷In addition, the absolute value of self-reporting error is symmetric around the norm. Self-reporting error as a *percentage* of the true weight is also increasing in absolute value given the absolute difference between the true weight and the norm, but the values are not symmetric around the norm.

The cost of reporting a weight that deviates from the truth, described above as $-\alpha(w^{SR} - w^T)^2$, can be seen as embedding two types of costs related to dishonesty: the intrinsic costs that reflect an internalized preference for honesty, and the extrinsic costs that arise if the surveyor can judge the honesty of the response. Specifically, the term α can be viewed as the sum of an intrinsic cost parameter, α^I , and an extrinsic cost parameter, α^E .

In the framework of Leary (1995), the extrinsic cost factor is referred to as “accountability.” For example, in a phone survey, self-reported values of an individual’s body weight will be subject only to intrinsic dishonesty costs. In an in-person survey, self-reported values will be subject to intrinsic costs as well as to extrinsic costs—to the extent that the interviewee believes that the interviewer might be able to detect, if imperfectly, a dishonest self-reported body weight value, especially if the self-reporting error is large.⁸ Therefore, for a given intrinsic cost schedule, the total costs of being dishonest are higher for an in-person survey, and therefore self-reports of body weight collected in person should be closer to the truth on average than those reports collected by phone or online. This prediction agrees with the findings of Pinkston (2015) and Courtemanche et al. (2015) described above, showing that survey mode affects the degree of bias in self-reports of weight and height.⁹

In the case of a weight examination that is preceded by conducting an in-person survey, as in the case of NHANES, the anticipated extrinsic dishonesty cost might be even higher. However, the subjects are told that the examiner will not be the same as the interviewer and that interviewers will not be able to observe subsequent exam weights among those that they interview.

3.3 Empirical Model

Equation (3) provides the structural foundation for the empirical analysis that follows. As noted above, the unrestricted NHANES data contain, for a consistent sample of people, self-reported values of body weight (and height) as well as examined values of body weight (and height). Self-reported values are obtained from in-person interviews and examined values of weight and height are measured at a mobile examination center (MEC), using a standard doctor’s scale and a height rod, respectively. The

⁸When survey responses cannot be validated based on visual inspection, such as in surveys of voting behavior, it has been observed that answers may be more accurate when the survey is self-administered (using paper forms or online interfaces) rather than administered by a live person, whether over the phone or in person. The reasoning is that social desirability bias stems from the desire to make a good impression on the surveyor, and this motive becomes more salient when the survey is administered by a person rather than via a more neutral instrument. Therefore, social desirability costs might also be modelled as depending on the survey mode.

⁹Confusing the matter, Cawley et al. (2015) attribute the greater accuracy of weight self-reports collected in person (as opposed to by phone) to social desirability bias. This conclusion makes sense only if social desirability bias is understood as the desire to avoid the embarrassment of getting caught in a lie. In the survey literature, however, social desirability bias is defined as responses being biased towards a socially desirable value for the behavior or outcome being measured, regardless of the truthfulness of the response. Therefore, we prefer to base the truth-telling motive on the costs of being dishonest.

examination takes place after the in-person interview, with a lag time that varies from zero to four months, with a mean lag of 25 days.¹⁰

In the empirical estimation, we define self-reporting error as the difference between the weight reported in the interview (self-reported weight) and the weight collected at the MEC (examined weight). This definition treats the examined weight as a proxy for the theoretical construct of true weight. In practice, of course, body weight is not a fixed value. Nonetheless, we assume that most individuals have a relatively stationary (i.e., nontrending) *average* weight value when considered over a period lasting from zero to four months, even though this weight will be subject to daily, weekly, or monthly fluctuations around this average. (The issue of time changes in body weight is discussed at length below.) Assuming also that the scales used by NHANES examiners are not systematically biased, the expected value of the examined weight equals this average value of the true weight. Therefore, the examined weight is assumed to measure an individual's true average weight, subject to classical measurement error.

For a hypothetical individual who (1) cares only about telling the truth, (2) has an accurate scale at home and uses it at fairly regular intervals such as once per month or more, and (3) has a body weight that is not trending upward or downward significantly over the relevant time frame, the self-reported weight should also conform (in expectation) to our practical definition of the individual's true weight. Under the stylized null hypothesis, then, the expected value of the self-reporting error is zero regardless of an individual's true weight, and therefore the regression coefficient on examined weight is expected to be zero.¹¹

This null hypothesis should hold for the hypothetical individual described above even if the home scale is less accurate than the scale at the mobile exam center (MEC), provided the direction of bias in the home scale is not correlated with an individual's true weight. Similarly, weighing conditions might differ between the home and the MEC—for instance, the amount of clothing an individual is wearing and the time of day that the weight measure is taken—and these variable factors could introduce differences between the self-reported weight values and the examined weight values even for a truthful individual who weighs herself regularly. At the MEC, individuals are typically weighed wearing only an examination gown and no shoes, but not at any particular time of day.¹² At home, weighing conditions will be less uniform. In sum, errors caused solely by differences in weighing conditions are not likely to be correlated with an individual's true weight.

¹⁰Using the NHANES data, the time lapse is calculated by subtracting age in months on the interview date from age in months on the date of the exam. Values for the difference are in whole months, ranging from zero to four. The average lapse of 25 days is based on an average lapse in months of 0.83 for women and 0.84 for men, multiplied by 30 days.

¹¹This statement holds, provided that fluctuations around the true (average) weight have a mean of zero regardless of the true weight, even if the fluctuations are heteroscedastic in true weight.

¹²For more information on weighing conditions at the MEC, see Cawley et al. 2015 and Zipf et al. 2013.

Nonetheless, to ensure that our results are not driven by moderate short-term weight fluctuations, differences in scale accuracy, and/or differences in weighing conditions, we set the weight self-reporting error to zero if the self-reported weight differs from the examined weight (on either side) by 2 kilograms (4.4 pounds) or less.¹³ If it has any effect, this modification should bias the results in favor of the null hypothesis just stated, since even modest self-reporting errors might be motivated by social desirability bias. In fact, the results are not significantly different without this modification.

The main test of the social desirability bias hypothesis involves testing that the coefficient on examined weight is significantly less than zero, but not less than -1 . As discussed below, there are a number of other reasons—aside from social desirability bias and mean-zero fluctuations—for why examined weight might differ systematically from self-reported weight. As we will show, however, it is hard to come up with a reasonable alternative story that yields the same qualitative predictions as our model. Therefore, a significant and robust negative coefficient (but not less than -1) on examined weight would provide strong support of our model’s validity.

The theoretical model above predicts that a linear relationship exists between an individual’s examined (true) weight and the self-reporting error. In the empirical analysis, we estimate the model as a linear spline in examined weight. This specification allows the slope of the line to vary over different portions of the examined/true weight distribution, and nests the strict linear relationship as a special case. (We discuss the choice of cut points for the splines below.) Within the conceptual framework, differences in the slope over different segments of the weight distribution are interpreted as differences in the marginal preference for truth-telling vs. norm-conformity in self-reporting behavior, although the slope should stay between zero and -1 over the entire range. In practice, the piecewise linear specification also facilitates tests of robustness to alternative explanations for misreporting, as discussed later in Section 6.

The model predicts that self-reporting error for body weight will vary positively with the assumed weight norm, where the slope on the weight norm will be opposite in sign (and equal in magnitude) to the slope on true body weight. In theory, the weight norm is assumed to vary only with sex and otherwise is held fixed across individuals. In practice, the weight norm is unobserved and may vary with other individual characteristics such as age and race. Given these facts, we include numerous controls that should help to proxy for the weight norm, as discussed below. To the extent that such controls are imperfect, the estimation could suffer from omitted variables bias. This will occur, for example, if the true body weight varies positively with the same norm that influences the self-reported weight (as in Burke and Heiland 2007). However, it can be shown that, under standard assumptions concerning the marginal effect of the norm on realized individual body weight, the omitted variables bias will run in a positive

¹³This approach follows Brestoff, Perry, and Van den Broeck (2011), who define self-reports as “accurate” if these differ absolutely from examined weight by two kilograms or less.

direction. With a positive bias, the estimated coefficient will be closer to zero, thereby reducing the chance that we obtain a significant negative coefficient.¹⁴

The basic empirical equation is as follows:

$$W_i^{SRE} = k + \vec{\alpha}\overline{W_i^X} + \beta\text{Height}_i + \gamma\text{Height}_i^2 + \delta\text{RoundError}_i + \theta\text{TimeLapse}_i + \vec{\mu}\overline{X}_i + \sigma\text{Survey}_i + \epsilon_i. \quad (5)$$

In the above, k is a constant term and W_i^{SRE} denotes the self-reporting error for body weight, defined as the difference between the individual's self-reported weight and her/his examined weight. The equation is not subscripted by time because we observe each person only once, even though not all individuals are observed at the same point in time. The term $\overline{W_i^X}$ refers to a vector of values, $\{W_i^{X1}, W_i^{X2}, W_i^{X3}, \dots, W_i^{Xmax}\}$, that together add up to the individual's examined weight, W_i^X . The value for the examined weight is partitioned in order to fit a piecewise linear function between specified nodes. The nodes for women are at zero pounds (node 0), 100 pounds (node 1), 200 pounds (node 2), and the maximum examined weight in the sample (443 pounds—node max).¹⁵ The term $\vec{\alpha}$ denotes a vector of coefficients, $\{\alpha^1, \alpha^2, \alpha^3\}$, each of which represents the slope of the line over the segment between the given node (indicated by the superscript) and the next-smallest node.

“Height” refers to an individual's examined height, which might influence self-reported weight even after controlling for examined weight. For the sake of simplicity, we abstract from height differences in the conceptual model, but it seems reasonable to expect the weight norms that influence self-reported values to vary with height. This would be the case if the weight norms reflect medical recommendations embodied in healthy ranges for BMI, or in weight-for-height charts that one might see in a doctor's office. However, we allow for nonlinear adjustment for height by including squared height as well (Height^2).

“RoundError” refers to the predicted self-reporting error that would arise if individuals round their (examined) weight up or down to the nearest value that is a multiple of five. For example, someone whose examined weight is 161 pounds has a predicted “focal” self-reported weight of 160 pounds and a

¹⁴The required assumption is that the marginal effect of the weight norm on equilibrium body weight is nonnegative and strictly less than one. In the social interactions literature, it is standard to assume an upper bound of one on analogous marginal effects. Burke and Heiland (2007) perform a calibration exercise that suggests that, for U.S. women, the relevant effect is in the neighborhood of 0.20, far from the required upper bound. Under this assumption, self-reported weight increases with the norm by a larger margin than actual weight does, and therefore the (signed) self-reporting error varies positively with the norm. In estimating equation (3), we have an omitted variable (the norm) that is positively correlated with an included variable (true weight) and positively correlated with the dependent variable (self-reporting error). Therefore the estimated coefficient on true weight will suffer from a positive bias.

¹⁵Given these nodes, the value of W_i^{X1} equals the lesser of examined weight (W_i^X) and 100 pounds; W_i^{X2} equals the lesser of ($W_i^X - 100$) and 100 pounds, but would-be negative values are set to zero; W_i^{X3} equals ($W_i^X - 200$), again bottoming out at zero. For example, an individual weighing 112 pounds has $W_i^{X1} = 100$, $W_i^{X2} = 12$, and $W_i^{X3} = 0$.

predicted self-reporting error of -1 pound; for someone weighing 202.5 pounds, predicted focal self-reported weight is 205 pounds and the predicted error is 2.5 pounds. We include this control based on evidence of focal-point rounding in other realms, as well as on the fact that the distribution of self-reported body weight in our sample has larger-than-expected mass on values ending in zero or five (see Figure 2, which shows a portion of the distribution of self-reported weights among women).¹⁶ RoundError thus represents an important control because focal point rounding might bias our tests for social desirability bias.¹⁷ The control only turns out to be a significant predictor of self-reporting error for women, but for consistency we include it also in the model for men.¹⁸

“TimeLapse” refers to the amount of time, expressed in months, between the date on which the individual had his/her weight and height examined and the date of the interview in which values for self-reported weight and self-reported height were collected. Given this time lapse—which in our sample averages 25 days and ranges from zero to four months—weight gain (or loss) during this interval may create the appearance of underreporting (or overreporting). The time lapse control will capture the average (signed) weight change per month in the population, ensuring that we do not attribute such changes to self-reporting error.¹⁹ Given the average tendency of individuals to gain weight as they age, we expect a (small) negative coefficient on the time lapse variable: the longer the time between the interview and the exam, the larger the weight gain, and therefore the smaller the value of the (signed) self-reporting error.²⁰

The term \vec{X}_i refers to a vector of individual characteristics that might affect the self-reporting error by influencing the weight norm that the individual considers most socially desirable. These include age, race/ethnicity, educational attainment, household income, smoking status, marital status, health insurance status, and disability status.

¹⁶This control is included even though we set self-reporting errors of 4.4 pounds or less to zero. It may nonetheless have explanatory power for larger errors that reflect, for example, rounding to the nearest 10-pound increment.

¹⁷The bias to the estimated coefficient could run in either direction. If individuals round up or down to the nearest multiple of five, then within each five-pound range of weight values the predicted slope of self-reporting error in true weight is negative, similar to the social desirability model. However, across the range of body weights, the errors predicted by focal-point rounding form a discontinuous function, with an average slope that might be either negative, positive, or zero. Also, self-reporting errors in this scenario should range only between -2.5 and 2.5 ; they will not get larger in absolute value at either very low or very high weight values, unlike in the social desirability model. The average value of self-reporting error under focal-point rounding may differ from zero if, for example, the true weight distribution is not uniform, or if either the minimum or maximum value of true weight is not a multiple of five.

¹⁸We also tested a model of biased rounding in which individuals round down to the nearest multiple of five, but this had weaker explanatory power (for women) than the neutral error model and no explanatory power for men.

¹⁹We exclude pregnant individuals from the analysis, so pregnancy-related changes in weight are not a concern.

²⁰Despite the average tendency of individuals to gain weight over time, some people may have been on a weight loss trajectory during the period of observation. Below we further discuss the robustness of our findings to the presence of individuals on significant weight loss or weight gain trajectories.

Finally, “Survey” refers to an indicator for the two-year NHANES cycle in which the observation was collected. There are a number of reasons why this factor might influence self-reporting errors for body weight. Changes in the aggregate weight distribution over time might cause a shift in the weight norm, according to the feedback process described theoretically in Burke and Heiland (2007). Evidence of a more forgiving weight norm in recent years is found in Burke, Heiland, and Nadler (2010) and Stommel and Osier (2013). In addition, the calibration of the scales used by NHANES examiners to measure body weight might vary at random over time across NHANES cycles, which, holding self-reported weight (and true weight) constant, would affect the measured value of self-reporting error..

4. Data and Sample Selection

Combining the six continuous NHANES waves, beginning with 1999–2000 and ending with 2009–2010, we construct one baseline sample for women and one baseline sample for men. The analysis is limited to observations from the examination sample for adults aged 20 years and older.²¹ The analysis is conducted separately for men and women. In order to eliminate influential outliers, we omit the top 0.25 percent of the gender-specific self-reporting error distribution within each sample.²² We retain those individuals with nonmissing values for the relevant controls. This last restriction does not significantly alter the mean value of self-reporting error in any sample.

In Table 1, the summary statistics refer to the distribution based on the entire 1999–2010 period.²³ The abbreviation “X” stands for “examined” and “SR” stands for “self-reported.” The average self-reporting error in weight is about –3 pounds for women and 0.04 pounds for men, suggesting that an average tendency to underreport weight is limited to women. The average self-reporting errors for height are positive (and relatively small) for both women and men, but the average is larger among men. The implied self-reporting error for BMI is negative for both women and men, but larger in absolute value for women.²⁴

²¹Examination weights are used, and we follow the methodology in the NHANES analytical guidelines for assigning weights when combining NHANES cycles. All analysis takes into account the complex survey design of the NHANES and uses the recommended methodology for analyzing subpopulations when relevant.

²²In cleaning the data we identified 1 or 2 extreme values for self-reported weight that we inferred as having been reported in kilograms but interpreted in pounds. We converted them to pounds.

²³In the table, “X Weight Neutral Rounding Error” refers to the “RoundError” variable defined above. The “All Hispanic” category includes anyone indicating Hispanic ethnicity regardless of other racial affiliations, and includes Mexican-Americans. Other racial/ethnic identities are defined so as not to overlap with Hispanic identity. The variables beginning with “PIR” refer to ranges of the Poverty Income Ratio, indicating household income relative to the threshold for official poverty status for a household of the same size.

²⁴These average signed errors reflect a cancelling out effect between significant positive and significant negative errors, as discussed in Cawley et al. (2015).

5. Results

5.1 Self-Reporting Errors in Body Weight as the Dependent Variable

5.1.1 Women

We begin by estimating equation (5) above on the NHANES 1999–2010 baseline sample for women. The results are presented in Table 2. The model in column 1 includes only the piecewise-linear terms for examined weight. The model in column 2 adds the list of controls described above. Additional controls are added in columns 3 and 4, and these results are discussed in Section 6.

Because the dependent variable is the signed self-reporting error, a positive coefficient means that the given factor predicts either an increase in the extent of overreporting of weight or a decrease in the extent of underreporting of weight; likewise a negative coefficient means that the factor predicts either a decrease in the extent of overreporting of weight or an increase in the extent of underreporting of weight. The estimated coefficient on “X Weight<100” refers to the marginal effect of examined weight on the self-reporting error in weight, for values of examined weight up to 100 pounds. The coefficient on “X Weight< 200” refers to the analogous marginal effect for values of examined weight between 100 and 200 pounds, and the coefficient on “X Weight>200” gives the slope for examined weight values 200 pounds and over. Among men (see Table 3), we use different cut points for the linear spline—for example the top group is 400 pounds and over—but the interpretation of coefficient estimates is analogous.²⁵

First considering columns 1 and 2, all the estimated coefficients on the “X Weight” variables lie strictly between 0 and –1 and are highly statistically significant. These results agree strongly with the prediction from the conceptual model. The fact that the estimates are closer to zero than they are to –1 indicates a modest degree of social desirability bias, in the sense that female respondents care more about reporting a truthful weight than they do about reporting a socially desirable weight.²⁶ Based on the specific point estimates for the different intervals, social desirability bias appears strongest among those with an examined weight of less than 100 pounds, and weakest for those with an examined weight over 200 pounds.²⁷ These slope estimates, when combined with the effects from remaining terms in the model, imply that self-reporting errors for body weight among women are positive in expectation for women with examined weight below 132 pounds, and negative in expectation for women weighing above 132

²⁵Different cut points were used for men to account for differences in the distribution of examined weight for men and women, as well as findings that a straight line offers a good fit to the male data for examined weights between 110 and 400 pounds.

²⁶A coefficient estimate of –0.5 on examined weight would indicate equally strong preferences for reporting a truthful weight value versus reporting a norm-conforming (i.e., socially desirable) weight value. Values between –0.5 and –1.0 would indicate a greater concern for reporting a weight that conforms to the norm rather than reporting a truthful value.

²⁷Additional tests (not shown) indicate that the coefficient estimates are not significantly different between the lowest weight interval and the intermediate interval, but that the coefficient estimates are significantly different between the intermediate and top intervals.

pounds (see Figure 3). This qualitative pattern agrees strongly with that predicted by the theoretical model and depicted in Figure 1. However, the agreement between the norms is merely coincidental.

In column 2, the significant positive coefficient on the “rounding error” variable indicates that self-reporting errors for women are also at least partly explained by rounding to the nearest five-pound increment. Being a current smoker (as opposed to a never-smoker) and being between 20 and 34 years of age (as opposed to being 45–59 years old) are factors both associated with significantly smaller (signed) self-reporting errors.²⁸ In contrast, being either Hispanic or black (as opposed to white), having a disability (versus having no disability), or having health insurance (versus having no insurance) all predict significantly larger self-reporting errors. As opposed to those surveyed in the 1999–2000 wave of NHANES, being surveyed in either the 2003–2004 wave, the 2005–2006 wave, or the 2009–2010 wave predicts a significantly larger error on self-reported body weight.

Perhaps surprisingly, the coefficient on examined height is not statistically significant (neither is the coefficient on height-squared), nor is the time lapse variable. However, the signs on these coefficients are as expected: being taller may be associated with somewhat higher self-reporting errors on weight (that is, less severe underreporting), and having a longer time lapse is associated with modestly lower self-reporting errors (that is, more severe underreporting), perhaps due to weight gain between the interview and the exam. Household income, based on our somewhat coarse measure, is not associated with significant differences in self-reporting errors among women, and among education categories only the “some college” group differs from those with just a high school diploma—the latter contrast is negative but only marginally statistically significant. The model’s explanatory power is modest, with an R-squared of 0.13, and most of this power comes from examined weight rather than demographic characteristics.

5.1.2 Men

Table 3 shows the corresponding results for men’s self-reporting errors in body weight as estimated over the NHANES 1999–2010 baseline sample. Again we consider only columns 1 and 2 for the time being. Among men, significant negative coefficients on the piecewise-linear terms in examined weight strongly agree with the presence of social desirability bias. Referring to the results shown in column 2, the coefficient on examined weight for values between 110 and 400 pounds is roughly equal to the coefficient on examined weight for women with weights between 100 and 200 pounds, suggesting the existence of a similarly modest degree of social desirability bias. The bias appears to be stronger for those individuals with examined weight values either below 110 pounds or over 400 pounds, and the difference in coefficients between the intermediate interval and the top interval is statistically significant. This latter result differs from the observation that social desirability bias for women appears to be weakest among

²⁸Throughout the discussion, “significant” differences are those that pertain to coefficient estimates that are statistically significant at the 5 percent level or better.

those with the highest examined weights. The steeper negative slope for men with extreme weight values may reflect a greater concern for reporting a socially desirable weight, or it may reflect other factors. Alternative explanations are discussed in Section 6 below. As seen in Figure 3 (which precedes Table 3), the predicted self-reporting errors for body weight among men are positive below a certain threshold (195 pounds) and negative above that threshold, consistent with the theoretical predictions.

As shown in Table 3, column 2, having a disability again predicts a larger (signed) self-reporting error. However, a number of the results presented in this column differ from those observed for women. For example, the time lapse carries a significant negative coefficient, indicating that each additional month of time between the interview and the exam reduces self-reported weight relative to examined weight by about one-fourth of a pound. Being a current smoker has a significant positive effect on the self-reporting error. Also, Hispanic men report lower weights (rather than higher) than whites for the same examined weight. Significantly smaller signed errors are also observed for those in the “Other/Multi” racial category, among never-married men (compared with currently married men), among men with some college (compared with those with only high school diplomas), and among men with bachelor’s degrees. The effects of the particular NHANES survey wave are never statistically significant.

5.2 Models with Height Self-Reporting Error as the Dependent Variable

The model of self-reporting behavior for body weight can be readily adapted to describe self-reporting behavior for height. Replacing the weight norm with a height norm, the model predicts that self-reporting errors in height will be positive for individuals with heights that fall below the height norm and negative for those with heights above the norm, resulting in a negative coefficient on examined height. The absolute value of the self-reporting errors should again be expected to increase in both directions as true height deviates more from the norm.

To test for social desirability bias in self-reporting errors for height, we adopt empirical models similar to those used for self-reporting errors in weight. The models are piecewise linear in examined height (rather than piecewise linear in examined weight), and include controls for examined weight and the square of examined weight.²⁹ To replace the rounding error variable used in the weight self-reporting model, we construct a rounding error variable for height that assumes individuals round up or down to the nearest whole inch. We set self-reporting error to zero for errors of plus or minus two centimeters, again following Brestoff, Perry, and Van den Broeck (2011).³⁰ The remaining controls, including the time lapse

²⁹Cut points for the spline are determined by inspecting the data and testing for significant differences in the slope across regions that appear to have different slopes. If the coefficient estimates are not significantly different between adjacent segments, the relevant cut point is dropped. The number of cut points has since been reduced relative to the model shown, but the results are qualitatively similar and the reduction in R-squared values is negligible.

³⁰Unlike body weight, height is not subject to daily or weekly fluctuations. However, it is difficult to accurately measure one’s height at home, and even a doctor may measure height with error.

variable, are identical to those used in the model of self-reporting errors for weight. In the model of height, we do not expect to observe a significant coefficient on the time lapse variable. Even though height tends to decline over time beginning at age 40, age controls should capture these differences. In any event, age-related declines between the interview and the exam are likely to be too small to measure, based on findings that after 40 years, individuals lose on average just 0.4 inches, or about one centimeter, per decade.³¹

Table 4 shows results from the model just described, using the NHANES 1999–2010 baseline samples for women (columns 1 and 2) and men (columns 3 and 4). For women with examined height values up to 64 inches—which is roughly equal to average female height in our sample as well as in the U.S. as a whole as of 2010—the coefficient on examined height is negative as predicted and height is overreported in expectation. Against the model of social desirability bias, the coefficient on examined height becomes positive for examined height values above 64 inches, and in expectation all women overreport their height by at least a small margin. (The positive coefficient on examined height is statistically significant only for examined height values between 64 and 67 inches, and becomes indistinguishable from zero for examined heights over 67 inches.)

The results just described do not line up fully with the predictions of the conceptual model, in the sense that we do not identify a height threshold or norm above which women underreport their height in expectation. However, the self-reporting errors predicted by the model (for an individual with average values for the controls) are equal to 0.5 inch or less for examined height values over 59 inches, as seen in Figure 4. This figure also shows that women with a stature of 57 inches or less overreport their height by 1 inch or more on average, and this overreporting of height becomes more severe as height declines below this threshold. These findings suggest that being taller is socially desirable, but being taller than average is not socially undesirable among women. In addition, women (as well as men) may be more apt to report height (rather than weight) truthfully because one has little if any control over one’s height. Among women, a number of significant negative health effects—including increased risk of ovarian cancer—have been associated with being very tall, but not with being very short. In light of this fact, the observed self-reporting patterns for height among women appear to be driven more by the social desirability of different statures than by the health consequences.

Among men (columns 3 and 4 of Table 4), the estimated coefficients on examined height are all significantly negative, but the absolute value of the coefficient is small for height values between 66 and 74 inches. As seen in Figure 4, predicted self-reporting errors for height are roughly 1 inch or greater for

³¹See, for example, Nagourney, E. 2013. “Why Am I Shrinking?” *The New York Times*, Booming Blog, Jan. 3, 2013. http://www.nytimes.com/2013/01/03/booming/why-am-i-getting-shorter-with-age.html?_r=0.

men with examined height values below 60 inches, declining to roughly 0.5 inch for a man of 66 inches. For examined heights between 66 and 74 inches, expected self-reporting errors are roughly flat in examined height, staying close to 0.5 inch. Above 74 inches, self-reporting errors decline more steeply and become negative for examined heights above 78 inches. While these findings are more consistent with the predictions of the model of social desirability bias, the results indicate a height norm for men of 78 inches. Again it appears that being taller is socially desirable at almost all heights and that being very tall is not socially undesirable unless one is exceptionally tall.

Other patterns in height self-reporting are worth noting because these differ in several respects from observed patterns in weight self-reporting. The neutral rounding error term is not significant for women, and carries a significant negative coefficient among men, indicating that men round their height up rather than rounding to the nearest inch in either direction. The coefficient on the time lapse variable is never significantly different from zero and has smaller magnitudes than it did in the models of weight self-reporting, consistent with what we would expect because—controlling for age effects and setting small errors to zero—height should not be trending in either direction. Compared with middle-aged women and men (45–59 years old), women and men/adults above age 60 overreport their heights by larger margins. This fact suggests that older subjects may not be aware of age-related declines in stature.

We observe no significant differences in self-reporting errors for women or men across the different NHANES survey cycles. In the case of body weight, the self-reporting errors among women appeared to increase (become less negative) in the later surveys. The contrast (among women) between the temporal patterns in weight reporting and height reporting is consistent with the proposition that the distributions of true height and weight influence norms and therefore self-reporting behavior: while the distribution of measured weight among women shifted to the right over this period, the distribution of measured height did not change significantly.

For both women and men, R-squared values for the models of height self-reporting errors are significantly lower than corresponding R-squared values for weight self-reporting errors, suggesting that the former are less amenable to explanation within our conceptual framework. The low explanatory power of the height models also reflects the fact that we set absolute reporting errors less than or equal to a range of zero to two centimeters, which suppresses variation in the dependent variable.

5.3 Models with BMI Self-Reporting Error as the Dependent Variable

Individuals do not self-report a BMI value directly in the NHANES interview, but we infer self-reported BMI using an individual's self-reported weight and self-reported height. As such, an individual may not be aware of his/her implied self-reported BMI value. Nonetheless it is important to examine self-reporting behavior for BMI because these implied values are used in the United States to calculate state-level and county-level obesity rates. Using models similar to those employed for height and weight above,

we test for social desirability bias in self-reporting of BMI, which will be indicated by significant negative coefficients on examined BMI that should not fall below -1 . We adopt a piecewise-linear spline function for examined BMI, with BMI cut points at 18.5 and 35.³² Regressions are conducted on the baseline samples described above.

Table 5 shows the results for women and Table 6 shows the corresponding results for men. In columns 1 and 2 of Table 5, the coefficients on all segments of examined BMI are consistent in sign and magnitude with the presence of social desirability bias, and are highly statistically significant for BMI values of 18.5 or greater. The coefficient on examined BMI for values below 18.5 becomes marginally significant when the basic controls are added, and becomes highly significant in columns 3 and 4 (discussed in Section 6) when additional controls are added. Similar to the results for weight self-reporting errors, the extent of social desirability bias is modest in the sense that the coefficients are all less than 0.5 in absolute value.

Patterns in self-reporting errors for BMI naturally reflect the influence of reporting patterns in both weight and height. Women aged 60–74 years, as well as those 75 years and over, have significantly smaller self-reporting errors for BMI than women aged 45–59 years, implying smaller overreporting margins or larger underreporting margins. These effects largely reflect the fact that older women are more likely to overreport their height, not that they are more likely to underreport their weight. Women in the highest household income category have larger self-reporting errors than those with lower incomes, suggesting a reduced tendency to underreport BMI; this fact derives primarily from the fact that higher-income women overreport their height by smaller margins than do lower-income women. Other significant effects for women follow the results pertaining to weight self-reporting errors. For example, BMI self-reporting errors are significantly greater among black women than among white women, and greater among women with health insurance. However, the survey cycle effects on self-reporting errors for BMI are mostly insignificant, despite the fact that weight reporting errors appeared to increase among women in later surveys.

The results for men (Table 6, columns 1 and 2) are broadly similar to those observed for women. The coefficients on examined BMI are negative and highly significant for all intervals of BMI, consistent with the presence of modest social desirability bias. For BMI values between 18.5 and 35, the coefficient

³²A BMI of 18.5 represents the lowest BMI value that is classified as “normal” (rather than “underweight”) by the CDC, and also corresponds roughly to the average BMI in our sample among those at or near the lower weight cut point of 100 pounds (for women) or 110 pounds (for men) used in the weight regressions. A BMI of 35 counts as obese (a BMI of 30 or greater is considered obese) in the CDC classification system and also represents the minimum BMI value for the CDC’s “class 2 obesity” category; in addition it is roughly the BMI equivalent of the upper weight cut point for women (200 pounds) used in the previous regressions. For men, 400 pounds (the upper cut point from the weight regressions) corresponds to a BMI over 50; a BMI cut point at 50 is not included for men because the slope on examined BMI above 50 was found not to differ significantly from the slope for the BMI interval between 35 and 50.

magnitudes are similar to those observed for women. While coefficient estimates—and therefore the estimated degree of social desirability bias—are more sharply negative among men with BMI values either below 18.5 or greater than 35, testing indicates that the slope in either of these regions does not actually differ significantly from the slope in the intermediate region of BMI.

Among men, a longer time lapse predicts a smaller BMI self-reporting error (more underreporting); this result suggests that the average man gained weight between the interview and the exam. However the effect size is small and significant only at the 10 percent level. Like older women, older men also underreport BMI more than middle-aged men, again based on their greater tendency to overreport their height. The effects of having either some college or a college degree are negative and statistically significant but not very economically meaningful because these effects are small in magnitude.

Based on these regression results, Figure 5 shows the fitted values of BMI self-reporting errors for women and men, respectively.³³ The control variables are held at their (gender-specific) weighted sample means. Consistent with the conceptual framework, the expected self-reporting errors are positive for examined BMI values below a threshold value (22 for women and 26 for men) and negative for examined BMI values above the threshold. Although not necessary based on the conceptual framework, the absolute self-reporting bias is greater on average among those who underreport their BMI than among those who overreport their BMI. This difference derives in part from the fact that a larger share of individuals underreport their BMI rather than overreport it.

5.4 Inferred Social Norms of Weight and BMI

In the conceptual model, individuals weighing less than a given value of body weight (construed as the social norm) will exhibit positive self-reporting errors, while those weighing more than the normative value will exhibit negative self-reporting errors. Based on the regression analysis, we can estimate global social norms for body weight and BMI, separately for women and men. We define the estimated norm to be the value of examined body weight (or examined BMI) such that, holding other characteristics at their sample means, the predicted value of self-reporting error for body weight (or BMI) equals zero.

Refer again to Figure 3, which shows the predicted values of weight self-reporting errors as a function of examined weight. The predictions are based on the regression models in Tables 2 and 3 (column 3 in each case), and hold all the control variables at their respective weighted (gender-specific) sample means. Among women, the inferred weight norm, at which an individual will truthfully report her weight, equals 132 pounds. For a woman of average height, which in our female baseline 12-year sample is roughly 64 inches, this inferred weight norm corresponds to a BMI of 22.7. This BMI value lies squarely within the “normal” BMI range of 18.5–24.9 promoted by the U.S. Centers for Disease Control (CDC), and falls 20

³³Fitted value plots are based on the regression models in Table 5 (column 3) for women and Table 6 (column 3) for men. The plots are not substantively different when based on the models in column 2 of either table.

percent below our sample mean BMI value among women. Among men, the inferred social norm for weight at 195 pounds, is greater. For a male of average height in our baseline sample (69 inches), this threshold corresponds to a BMI value of 28.8. This figure is roughly equal to the male sample mean BMI of 28.4, but nonetheless falls within the “overweight” BMI category (25-29.9) set by the CDC. However, these inferred BMI norms do not incorporate self-reporting errors for height.

Referring again to Figure 5, for women the inferred social norm for BMI equals 22, slightly below the estimate of 22.7 that was based only on self-reporting errors for weight. For men, the inferred social norm for BMI equals 26, substantively lower than the previous estimate of 28.8 and just above the official “overweight” threshold of 25. This lower norm estimate reflects the impact of height overreporting, which on average is more severe among men than it is among women.

If these latter thresholds are taken to represent social norms for BMI, they indicate that women refer to a lower BMI norm (22) than men (26), and only the female norm lies in the normal range set by the CDC’s public health standards. The gender gap between the thresholds accords with evidence from numerous studies (such as Bordo 2003 and Leit, Gray, and Pope 2001) suggesting that the perceived ideal physique among women is quite thin, while the ideal physique for men is larger and emphasizes muscularity. Because muscle weighs more than fat, a muscular male norm might contribute to a normative BMI that exceeds the overweight threshold but that nonetheless does not visually present itself as overweight. However, it may also be the case that a certain amount of body fat is more acceptable among men than among women.

6. Robustness Analysis: Can Time Trends and Infrequent Weighing Explain Our Results?

There are several reasons other than social desirability bias for why self-reported values of weight might differ from examined values. We have already controlled for factors such as focal-point rounding, random short-term weight fluctuations, differences in scale accuracy and weighing conditions between an individual’s home and the MEC, and average short-term weight trends. In this section we address objections concerning the relevance of weight norms as a motivation for self-reporting errors, and assess the validity of alternative explanations for self-reporting errors based on heterogeneous time trends in weight and an individual’s lack of awareness of body weight due to infrequent weighing. Considered in combination with each other, we find that these factors may contribute to severe weight underreporting among very heavy individuals, giving the appearance that social desirability bias is especially strong among that group. At the same time, evidence of social desirability bias is found to be largely robust even among those who are most likely to be aware of their current weight and allowing for heterogeneous short-term time trends in weight.

6.1 Testing the Relevance of Norms

Brestoff, Perry, and Van den Broeck (2011), using Irish survey data from 2007, question the role of social desirability bias in explaining self-reporting errors for body weight. Controlling for objective weight status, they find that those who describe themselves as “too heavy” based on a direct question are less likely to underestimate their BMI than those who describe themselves as being “about the right weight.” They argue that those who describe themselves as “too heavy” are more likely to be aware of social norms concerning ideal body weight standards and that if these norms motivate misreporting, such individuals should be more likely, not less likely, to underreport their weight or BMI. Instead, underreporting of weight among overweight subjects who describe their weight as “about right” may indicate that such individuals weigh themselves (or get weighed) infrequently and have gained weight since their last time on a scale.

The finding that arguably more “norm-aware” individuals are less likely to understate their BMI does not imply that such individuals never understate their BMI. Social desirability bias may still prevail among this group, while infrequent weighing may drive misreporting among those who are arguably less aware of social norms and/or less aware of their own true weight. To test these possibilities, we use NHANES data pertaining to a similar question—whether an individual considers himself (at his current weight) to be “overweight,” “about right,” or “underweight.” We similarly define “norm-aware” individuals as those who consider themselves to be “overweight.” This is admittedly a loose proxy for norm awareness, since those who describe themselves as “about right” might also be norm aware and happen to be neither overweight nor underweight; conversely those who describe themselves as “overweight” might not actually be overweight but responding to a distorted norm. Those self-classifying as “underweight” might also be norm-aware, and we control for this as well.

If a lack of awareness of norms is associated with having large self-reporting errors, then a failure to control for norm awareness (or the lack thereof) might bias our results in favor of finding a negative relationship between examined weight and the self-reporting error for weight. To guard against this possibility, we rerun the basic regressions from Tables 2 and 3, adding a binary indicator for a weight perception of “about right” to the model, but using the same sample. We also add an indicator for a perception of “underweight” so that the omitted category includes only those who consider themselves overweight. The results are given in column 3 of Table 2 (women) and column 3 of Table 3 (men).

Among both women and men, those who perceive their weight as being “about right” exhibit significantly smaller (signed) self-reporting errors than otherwise similar individuals (with the same examined weight) who perceive themselves as being overweight. This finding agrees with the results in Brestoff, Perry, and Van den Broeck that those individuals who are less aware of social norms are more likely to underreport their weight and/or to underreport it by a larger margin than those who are more

aware of norms.³⁴ However, the coefficients on the examined weight variables are not significantly different from those shown in column 2; this result offers preliminary evidence that the negative relationship between examined weight and self-reporting errors is not driven solely by those who are unaware of social norms. In subsection 6.3, we run regressions on subsamples defined by norm awareness as an additional robustness test.

6.2 Controlling for Heterogeneous Short-Term Trends in Body Weight

As discussed above, an individual might experience a significant weight gain or loss during the time between the interview and the exam. Including the time lapse as a control variable in the regressions captures how the average (signed) weight change per month in the population may contribute to apparent self-reporting errors. However, the linear time lapse control will be insufficient if weight gains and losses (per month) vary systematically with an individual's examined weight. For example, if those individuals with the largest (smallest) examined weights are those who experienced the largest recent weight gains (losses), it will be as if those with the largest (smallest) examined weights underreport (overreport) their weight by the largest margins, and self-reporting errors will be negatively associated with examined weight. These are not implausible propositions because starting from the same initial weight, those who gain more (lose more) will have a higher (lower) final weight, and the examined weight captures the final weight value.

To more flexibly control for weight changes between the interview and the exam, we include interaction terms between the time lapse variable and the binary indicator variables for being in a given interval of the distribution of examined weight. We create one indicator for each of the intervals defined by the cut points of the spline in examined weight, less one indicator for one of the intermediate intervals. For example, “[X Weight<100] x Time Lapse” represents an interaction between the binary indicator for having an examined weight under 100 pounds and the time lapse. These terms allow the effect of the time lapse to vary depending on the weight interval, which allows for differences in average weight changes per unit of time for people in different portions of the examined weight distribution.

The results of the regressions that include the interaction terms (between the time lapse and weight interval dummies) are shown in column 4 of Tables 2 and 3. Among women (Table 2, column 4), coefficients on examined weight are not significantly different than in the previous models (column 2 and 3) that included only a linear time lapse control. In addition, the effect of the time lapse does not differ significantly for individuals with either very low weights (below 100 pounds) or very high weights (over

³⁴As seen in column 3 of Table 3, those who classify themselves as underweight also exhibit significantly smaller self-reporting errors. In most cases, those who perceive themselves as underweight are objectively underweight, and the objectively underweight tend to overstate their weight. Therefore the negative coefficient on underweight perception indicates that, among individuals who are truly underweight, those who perceive that they are underweight are less likely to overreport their weight—that is, they are likely to have more accurate self-reports.

200 pounds) than among those with intermediate weight values. Therefore, among women, heterogeneous short-term time trends are not apparent and do not weaken the evidence supporting the presence of social desirability bias.

Among men (Table 3, column 4), adding the interaction terms results in a somewhat smaller absolute coefficient on examined weight for those below 110 pounds; in addition, the statistical significance of that coefficient is weakened from the 5 percent level to the 10 percent level. The coefficients on examined weight above 110 pounds are robust. These results suggest that men with examined weights below 110 pounds may have lost weight in the time between the interview and the exam, although the coefficient on the interaction term is not statistically significant. If weight losses did indeed occur, these could help to explain why members of this low-weight group tend to have positive self-reporting errors for weight. Therefore, the case for social desirability bias is somewhat weakened among low-weight men, but not completely eliminated, but it is highly robust among men weighing over 110 pounds.

6.3 Controlling for Infrequent Weighing and Longer-Term Weight Changes

The flexible controls for the contribution of the time lapse to self-reporting errors will be insufficient, however, if reporting errors reflect a combination of very infrequent weighing and longer-term trends in body weight. If self-reported weights are several months or even years out-of-date, the time lapse will be a poor proxy for the interval since the individual last weighed himself. Infrequent weighing, combined with a significant weight gain or loss since the last weigh-in, might produce weight self-reporting errors that are negatively correlated with the examined weight, thus spuriously creating the appearance of social desirability bias.³⁵

To create this effect, a sufficient condition is that those with larger examined weights experienced larger weight gains since their last weigh-in, and those with smaller examined weights experienced larger weight losses. As stated above, experiencing a large weight gain (or a large loss) will tend to place one in the group of heavier (lighter) individuals, and therefore this condition is not implausible. Against the outcome, however, those with large weight gains might have had very low initial weights, and those with large weight losses might have had very high initial weights, such that we might observe no correlation (or even the opposite relationship) between final weights and weight changes.

The NHANES lacks direct information on when an individual was last weighed. To proxy for weighing frequency or an awareness of one's own current weight, we use two separate pieces of information. The first is the same one used to proxy for an awareness of weight norms—an individual's

³⁵The results in DeAndrea et al. (2012) suggest that by itself, an individual's lack of awareness of her/his weight—even absent weight changes over time—may result in self-reports of weight that exhibit social desirability bias. The authors attribute this outcome to unconscious but self-serving bias in the face of uncertain knowledge. While not deliberate, this source of social desirability bias nonetheless reflects an innate awareness of weight norms and a desire to appear to conform to them.

subjective perception of their weight. Those who perceive that they are “overweight” will be considered more aware of their current weight, whereas those who perceive that they are “about right” will be considered less aware of their current weight.³⁶ For each sex, we split the baseline sample into these two groups and use the resulting subsamples to run regressions that are otherwise similar to those already described. The second proxy is derived from questions about recent attempts to control one’s weight. Those who indicated in the NHANES interview that they “tried to lose weight” in the past year, or that they “tried not to gain weight” in the past year, are more likely to have weighed themselves recently than those who did not try either to lose weight or maintain their weight. For each sex, we run the regressions on the subsample consisting only of those who tried to lose or not gain weight.

If it is merely an individual’s lack of awareness of his/her actual weight—combined with weight changes—that explains the negative relationship between self-reporting errors and examined weight, we should expect to observe this relationship only among those who are unaware of their current weight—as proxied by the “about right” perception. If social desirability bias drives the self-reporting errors, then the negative relationship should also be found among those who are aware of their current weight, as proxied either by the “overweight” perception or the “tried to lose/not gain” condition.

6.3.1. Subsample Regression Results for Women

The results for weight self-reporting errors for women are given in Table 7. For each subsample, we run two models: one that contains the standard controls and one that also includes interactions between the time lapse and the weight interval indicators. Columns 1 and 2 show the results for the subsample of women who perceive that they are overweight, not including members of this group with examined weight values of less than 100 pounds. Columns 3 and 4 show the results for the subsample of women who tried to lose weight or not gain weight in the past year. The results presented in these four columns therefore reflect behavior among women who are arguably aware of their current weight, based on one of two different proxies for such awareness. In columns 1 and 2 as well as in columns 3 and 4, the coefficients on examined weight—ignoring weights under 100 pounds—are not significantly different than those observed in Table 3. The results also do not differ significantly depending on whether interactions between the time lapse and the weight interval dummies are included. Therefore patterns consistent with social desirability bias are robust among women over 100 pounds even when we limit the sample to those who (based on these proxies) are likely to be aware of their current body weight. The bias

³⁶Somewhat coarsely, we assign these classifications regardless of an individual’s objective weight status. However, we drop from the sample those objective weight categories that are grossly inconsistent with the given self-perception. For example, among those who perceive that they are overweight, we omit women whose examined weight is below 100 pounds, and men with examined weight below 110 pounds, and similarly omit the lowest BMI category in the BMI regressions.

is modest among those who weigh between 100 and 200 pounds, and milder for those over 200 pounds, especially based on the results presented in columns 3 and 4.³⁷

The results in columns 5 and 6 pertain to the subsample of women who perceive that they are “about right” and are thereby considered less aware of their weight. Interestingly, the coefficients on examined weight are more strongly negative for this group than for the more weight-aware sample, especially for those with examined weights above 200 pounds. Such women, unless extremely tall, are very likely to be obese—a woman of average height at this weight has a BMI of 34, well above the threshold of 30 that defines obesity. Therefore, women over 200 pounds who perceive their weight as being “about right” would appear either to be grossly unaware of their weight or to have a distorted perception of what constitutes being overweight. Among this group, the negative correlation between self-reporting errors and examined weight may reflect a lack of weight awareness combined with weight gain since their last weigh-in, rather than social desirability bias. However, as suggested by DeAndrea et al. (2012), underreporting by weight-unaware subjects may stem from an unconscious form of social desirability bias, whereby instead of selecting random values in the face of ignorance people select socially desirable values. This latter explanation cannot be verified, however, so our robustness assessments err on the side of the explanation based purely on weight unawareness, meaning the explanation does not refer to any bias related to social desirability.

The negative correlation between self-reporting errors and examined weight is especially sharp for those women over 200 pounds who are potentially unaware of their true weight. This result suggests that infrequent weighing combined with longer-term weight gain together might contribute to extreme weight underreporting among those with high examined weights. In contrast, the results pertaining to weight-aware individuals suggest that social desirability bias induces relatively modest weight underreporting at the margin among those over 200 pounds. However, the coefficient in the full sample, which combines the effects of more-aware and less-aware types, is much closer to (or not significantly different than) the results estimated over the weight-aware sample. This finding suggests that the share of women over 200 pounds who perceive their weight to be about right is small enough so that the impact of infrequent weighing on the aggregate effects of self-reporting errors is small.

Figure 6 depicts the predicted self-reporting errors for weight based on the regressions just described, conducted on the restricted subsamples of women who are classified as “weight-aware” and “weight-unaware.” For comparative purposes, the fitted values for the baseline sample are repeated here. As

³⁷For women who are less than 100 pounds, the coefficient on examined weight is only marginally significant in column 3 and becomes small (positive) and insignificant in column 4. These results suggest that social desirability bias is not present among weight-aware women who weigh less than 100 pounds, although this is likely to be a small group based on how we proxy for weight awareness. (Based on their small numbers, these women are dropped from the sample in columns 1 and 2.)

discussed above, the behavior of weight-aware women (as proxied by “perceive overweight” or “tried to lose or not gain weight”) strongly mirrors behavior for the complete sample; however, the “about right” group exhibits a sharper negative relationship between examined weight and self-reporting errors, especially among women over 200 pounds. The inferred weight norm for this group is also lower, at 126 pounds rather than 132 pounds.

6.3.2. Subsample Regression Results for Men

Table 8 shows the results for the special subsamples for men. The first four columns represent the results for “weight-aware” individuals. In columns 1–4, those with weight below 110 pounds are dropped due to small sample sizes. In columns 1–3, the coefficients on examined weight are basically unchanged from those obtained over the full baseline sample (shown in Table 3, columns 2–4), indicating that weight-aware individuals offer robust evidence of displaying at least modest social desirability bias. However, in column 4 the coefficient on examined weight for those over 400 pounds becomes smaller in absolute value and is only marginally significant. This change occurs when the interaction between the time lapse and the indicator for “weight over 400 pounds” is added to the model. The coefficient on that interaction is a large and highly significant negative value, which suggests that, among men who are over 400 pounds, sharp weight gains between the interview and the exam contribute to the fact that they underreport their weight by large margins. Nonetheless, evidence of social desirability bias is not eliminated altogether for this group and remains robust for men with more moderate weights (between 110 and 400 pounds).

Columns 5 and 6 of Table 8 show the results for men who may be unaware of their body weight, based on perceiving that their weight is “about right.” No men in this group had an examined weight over 400 pounds, so that category is dropped from the regression. The coefficient on examined weight for those weighing between 110 and 400 pounds is negative and significant, and has a larger absolute value than the corresponding coefficient for arguably “weight-aware” men in the same weight range. When interactions between the time lapse and the indicator for weighing less than 110 pound are included (column 6), the coefficient on examined weight (for men weighing under 110 pounds) becomes positive and statistically insignificant. The positive coefficient on the interaction term indicates that, among those with low examined weights, weight loss between the time of the interview and the exam contributes to overreporting of weight. Therefore, among low-weight men who perceive their weight to be “about right,” overreporting cannot be linked to social desirability bias but is better accounted for on the basis of recent weight loss.

Figure 7 shows the predicted self-reporting errors among the restricted subsamples of men, based on the results presented in Table 8. As was the case for women, men that perceive their weight to be “about right” exhibit a steeper negative relationship between examined weight and the self-reporting error for

weight. Also, the steep slope on examined weight for those with weighing under 110 pounds is seen to be driven by the weight-unaware (“about right”) sample, while the steep slope among men over 400 pounds prevails in the weight-aware samples (but is unobserved in the weight-unaware sample).

7. Conclusion

Consistent with the predictions of our conceptual model, we find that individuals whose true body weight falls below a certain threshold tend to overstate their weight, while individuals above the threshold tend to understate their weight. Also consistent with the model, the degree of overstatement or understatement increases on average with the difference between the individual’s true weight and the threshold weight. Self-reporting errors for height behave somewhat differently compared to those related to weight. With the exception of extremely tall men, self-reported height is overstated on average at nearly all points in the distribution of measured height for both sexes. However, height self-reporting behavior exhibits social desirability bias in the sense that shorter individuals tend to overstate their height by a larger margin than do taller individuals. Also, above certain gender-specific height thresholds, self-reporting errors for height on average are small.

Self-reporting errors for BMI, which are inferred by combining self-reported values of weight and height, conform strongly to the predictions of the model’s conceptual framework, indicating that self-reports of BMI are influenced more by self-reports of weight than of height. Among women, expected self-reporting errors for BMI are positive for examined BMI values below 22, negative for BMI values above that threshold, and increase in absolute value away from the threshold in either direction. Among men, the pattern is qualitatively similar but the threshold, at a BMI of 26, is higher.

These empirical BMI thresholds that separate overreporting from underreporting—at 22 for women and 26 for men—can be interpreted as the most socially desirable values for BMI by gender. The value for women lies within the CDC’s recommended “normal/healthy” BMI range of 18.5-24.9, suggesting that public health standards contribute to norm formation among women. However, the gender difference between the thresholds suggests that norms contain at least some social component not readily explicable in terms of health standards. In particular, the higher BMI norm for men may reflect either a masculine ideal of muscularity, or a greater social acceptance of men who are overweight.

Among men, we find evidence that weight changes between the survey interview and the exam contribute to apparent self-reporting errors among men with either very high or very low body weights, but similar changes do not account for self-reporting errors for men over the bulk of the weight distribution. Among women, we find evidence that infrequent weighing combined with longer-term weight changes may contribute to underreporting of weight among very heavy women. In addition, biased weight reporting among those who are less aware of their true body weight may reflect an unconscious form of social desirability bias. However, these competing hypotheses for the behavior of arguably

“weight-unaware” individuals cannot be readily disentangled because the data lack direct information on weight self-awareness.

Our proxies for awareness of weight and awareness of social norms are admittedly imperfect. Therefore, one might claim that we cannot rule out the possibility that infrequent weighing combined with weight changes can explain most instances of self-reporting errors. However, previous research finds that the mode of data collection—for example, a face-to-face interview versus a phone interview—significantly alters the degree of bias in self-reported values of weight, even for the same individual. This evidence is inconsistent with explanations based on sheer ignorance of weight and lends credibility to our theoretical model. In the model, there is a tradeoff between social desirability bias and dishonesty costs, which are heightened when the weight self-report is subject to some degree of external validation, either by visual inspection or by an eventual examination.

To the extent that dishonesty costs work against social desirability bias, survey instruments should (and often do) prompt individuals to consider the benefits to researchers of providing honest reports. To the extent that ignorance of one’s true body weight contributes to reporting errors, individuals might be asked to weigh themselves in private on a home scale, and be provided with a scale if they do not have one, before providing a self-reported value of body weight. Our findings suggest, however, that this latter measure alone would not eliminate systematic biases rooted in social norms that affect self-reported values of body weight.

References

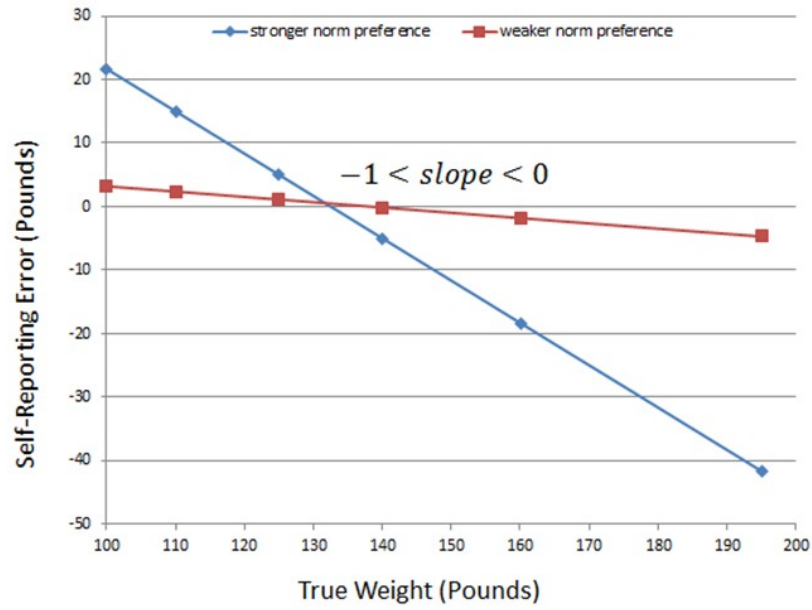
- Ambwani, Suman, and Jennifer F. Chmielewski. 2013. "Weighing the Evidence: Social Desirability, Eating Disorder Symptomatology, and Accuracy of Self-reported Body Weight Among Men and Women." *Sex Roles* 68(7): 474–483.
- Bordo, Susan. 2003. *Unbearable Weight: Feminism, Western Culture, and the Body*. 10th Anniversary Ed. Berkeley: University of California Press.
- Brestoff, Jonathan R., Ivan J. Perry, and Jan Van den Broeck. 2011. "Challenging the Role of Social Norms Regarding Body Weight as an Explanation for Weight, Height, and BMI Misreporting Biases: Development and Application of a New Approach to Examining Misreporting and Misclassification bias in surveys." *BMC Public Health* 11:331.
- Burke, Mary A., and Frank W. Heiland. 2007. "Social Dynamics of Obesity." *Economic Inquiry* 45(3): 571–591.
- Burke, Mary A., Frank W. Heiland, and Carl M. Nadler. 2010. "From 'Overweight' to 'About Right': Evidence of a Generational Shift in Body Weight Norms." *Obesity* 18(6): 1226–1234.
- Cawley, John. 2004. "The Impact of Obesity on Wages." *Journal of Human Resources*. 39(2): 451–474.
- Cawley, John, Johanna Catherine Maclean, Mette Hammer, and Neil Wintfeld. 2015. "Reporting Error in Weight and its Implications for Estimates of the Economic Consequences of Obesity." *Economics and Human Biology* 19: 27–44.
- Connor Gorber, Sarah., Mark S. Tremblay, D. Moher, and B. Gorber. 2007. "A Comparison of Direct vs. Self-Report Measures for Assessing Height, Weight, and Body Mass Index: A Systematic Review." *Obesity Reviews* 8(4): 307–326.
- Connor Gorber, Sarah, Margot Shields, Mark S. Tremblay, and Ian McDowell. 2008. "The Feasibility of Establishing Correction Factors to Adjust Self-Reported Estimates of Obesity." *Health Reports* 19(3): 71–82.
- Courtemanche, Charles, Joshua Pinkston, and Jay Stewart. 2015. "Adjusting Body Mass for Measurement Error with Invalid Validation Data." *Economics and Human Biology* 19: 275–293.
- DeAndrea, David C., Stephanie Tom Tong, Yuhua Jake Liang, Timothy R. Levine, and Joseph B. Walther. 2012. "When Do People Misrepresent Themselves to Others? The Effects of Social Desirability, Ground Truth, and Accountability on Deceptive Self-Presentations." *Journal of Communication* 62(3): 400–417.
- Engstrom, Janet L., Susan A. Paterson, Anastasia Doherty, Maryu Trabulsi, Kara L. Speer. 2003. "Accuracy of Self-Reported Height and Weight in Women: An Integrative Review of the Literature." *Journal of Midwifery and Women's Health* 48(5): 338–345.
- Gil, Joan, and Toni Mora. 2011. "The Determinants of Misreporting Weight and Height: The Role of Social Norms." *Economics and Human Biology* 9(1): 78–91.

- Groesz, Lisa M., Michael P. Levine, and Sarah K. Murnen. 2002. "The effect of experimental presentation of thin media images on body satisfaction: A meta-analytic review." *International Journal of Eating Disorders* 31(1): 1–16.
- Holbrook, Allyson L., and Jon A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74.1: 37–67.
- Karp, Jeffrey A., and David Brockington. 2005. "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries." *The Journal of Politics* 67(3): 825–840.
- Kelly, Nicole L. 2015. "Examining Social Desirability Bias in Measures of Financial Behavior." Theses and Dissertations. Paper 464. Available at <http://ir.library.illinoisstate.edu/cgi/viewcontent.cgi?article=1463&context=etd>.
- Krumpal, Ivar. 2013. "Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review." *Quality & Quantity* 47(4): 2025–2047.
- Leary, Mark R. 1996. *Self-Presentation: Impression Management and Interpersonal Behavior*. Boulder, CO: Westview Press.
- Leit, Richard A., Harrison G. Pope Jr., and James J. Gray. 2001. "Cultural Expectations of Muscularity in Men: The Evolution of Playgirl Centerfolds." *International Journal of Eating Disorders* 29(1): 90–93.
- Leit, Richard A., James J. Gray, and Harrison G. Pope Jr. 2002. "The Media's Representation of the Ideal male body: A cause for muscle dysmorphia?" *International Journal of Eating Disorders* 31(3): 334–338.
- National Institutes of Health. 1998. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report*. NIH Publication No. 98-4083. Bethesda, MD: National Heart, Lung, and Blood Institute. Available at https://www.nhlbi.nih.gov/files/docs/guidelines/ob_gdlns.pdf.
- Pinkston, J.C. 2015. "The Dynamic Effects of Obesity on the Wages of Young Workers." University of Louisville, Mimeo. Available at SSRN: <http://ssrn.com/abstract=2537554>.
- Rowland, M. L. 1990. "Self-Reported Weight and Height." *American Journal of Clinical Nutrition* 52(6): 1125–1133.
- Stommel, Manfred, and Charlotte A. Schoenborn. 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(1): 421.
- Stommel, M., and N. Osier. 2013. "Temporal Changes in Bias of Body Mass Index Scores Based on Self-Reported Height and Weight." *International Journal of Obesity* 37(3): 461–467.
- Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5): 859–883.
- van de Mortel, Thea F. 2008. "'Faking It': Social Desirability Response Bias in Self-Report Research." *Australian Journal of Advanced Nursing* 25(4): 40–48.

Yun, S., B-P. Zhu, W. Black, and R.C. Brownson. 2006. "A Comparison of National Estimates of Obesity Prevalence from the Behavioral Risk Factor Surveillance System and the National Health and Nutrition Examination Survey." *International Journal of Obesity* 30(1): 164–170.

Zipf, George, Michele Chiappa, Kathryn S. Porter, Yechiam Ostchega, Brenda G. Lewis, and Jennifer Dostal. 2013. "National Health and Nutrition Examination Survey: Plan and Operations, 1999–2010." National Center for Health Statistics. Vital and Health Statistics Series 1, Number 56.

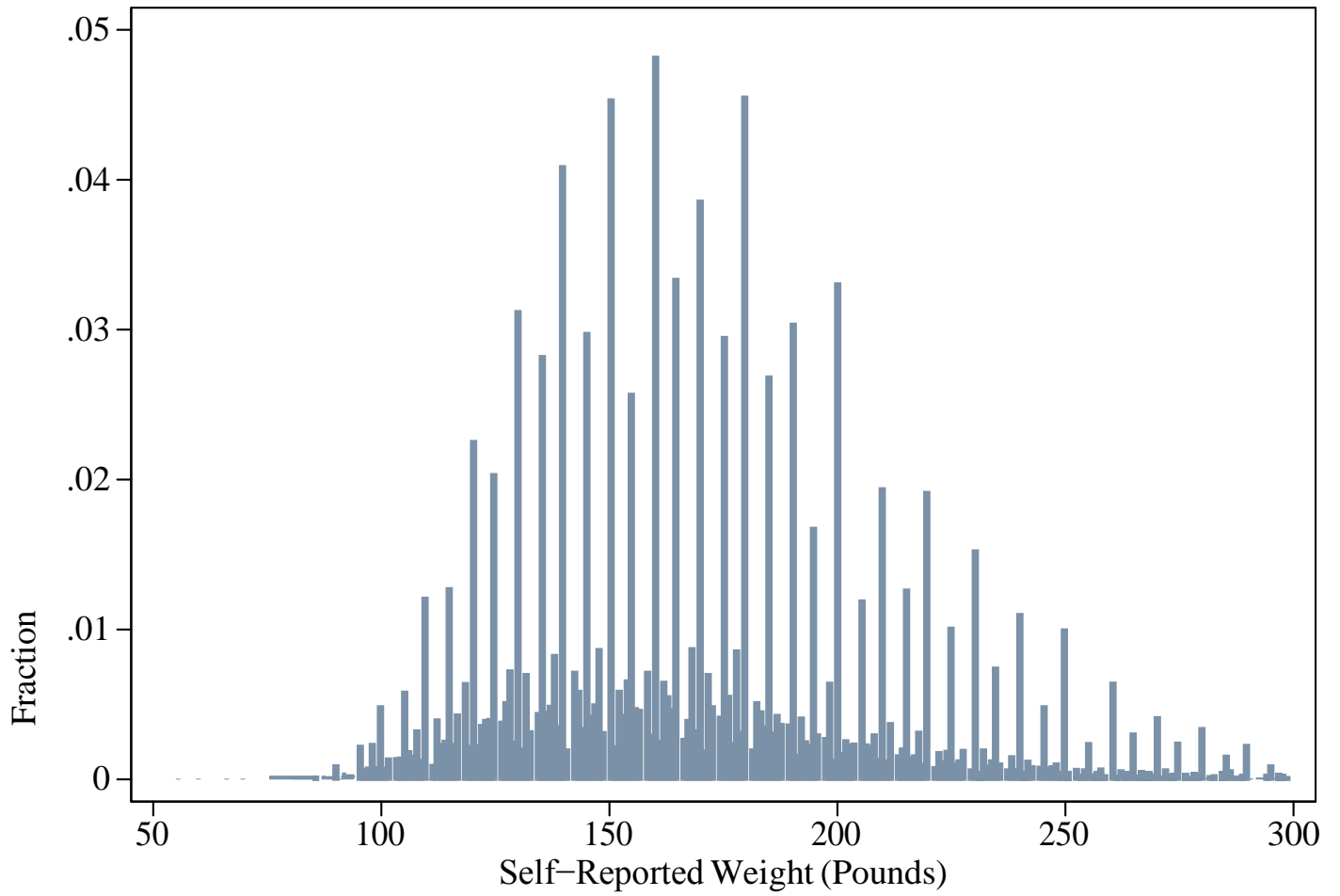
Figure 1: Predicted Self-Reporting Errors as Function of True Weight



Source: Authors' calculations based on hypothetical data.

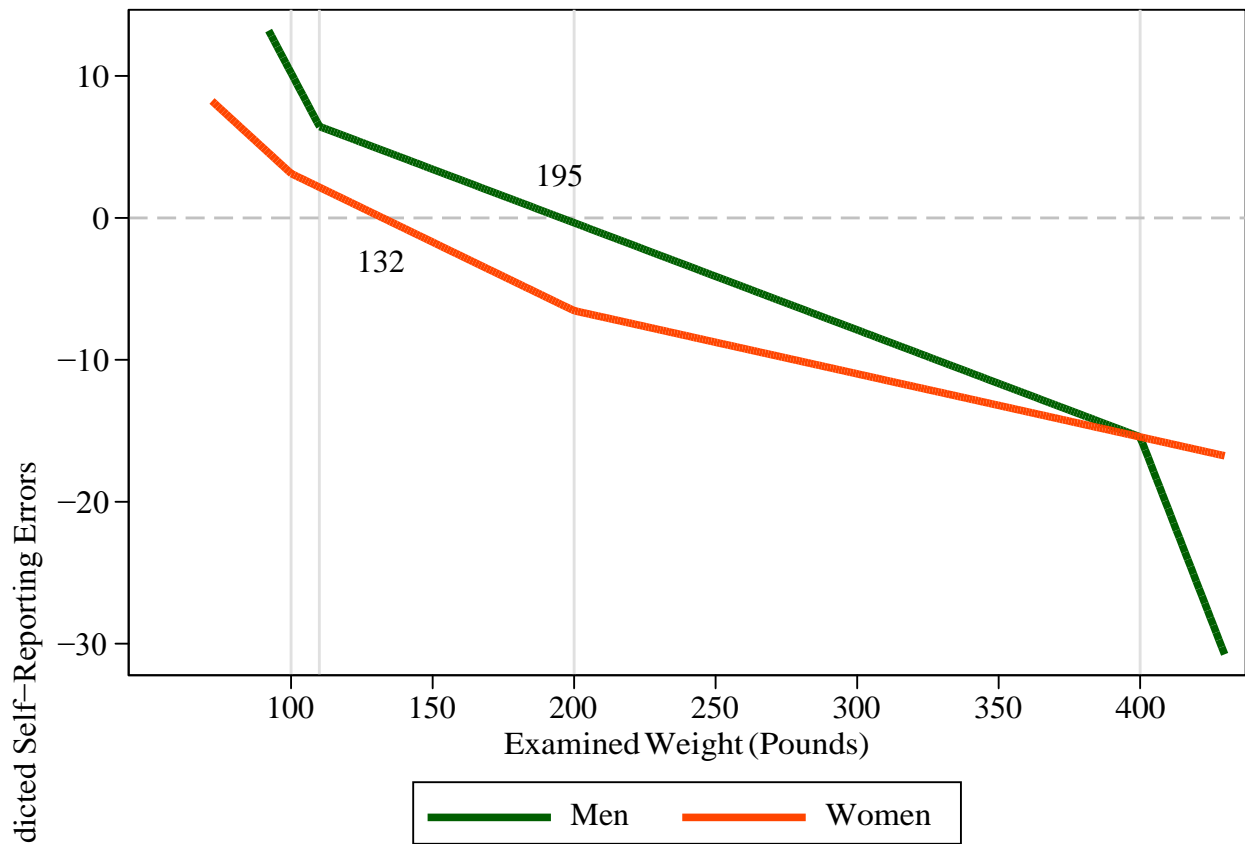
Note: The weight norm has been set at 132 pounds. The illustration represents a hypothetical scenario for women.

Figure 2: Women's Self-Reported Weight Clusters at Five- and Ten-Pound Increments



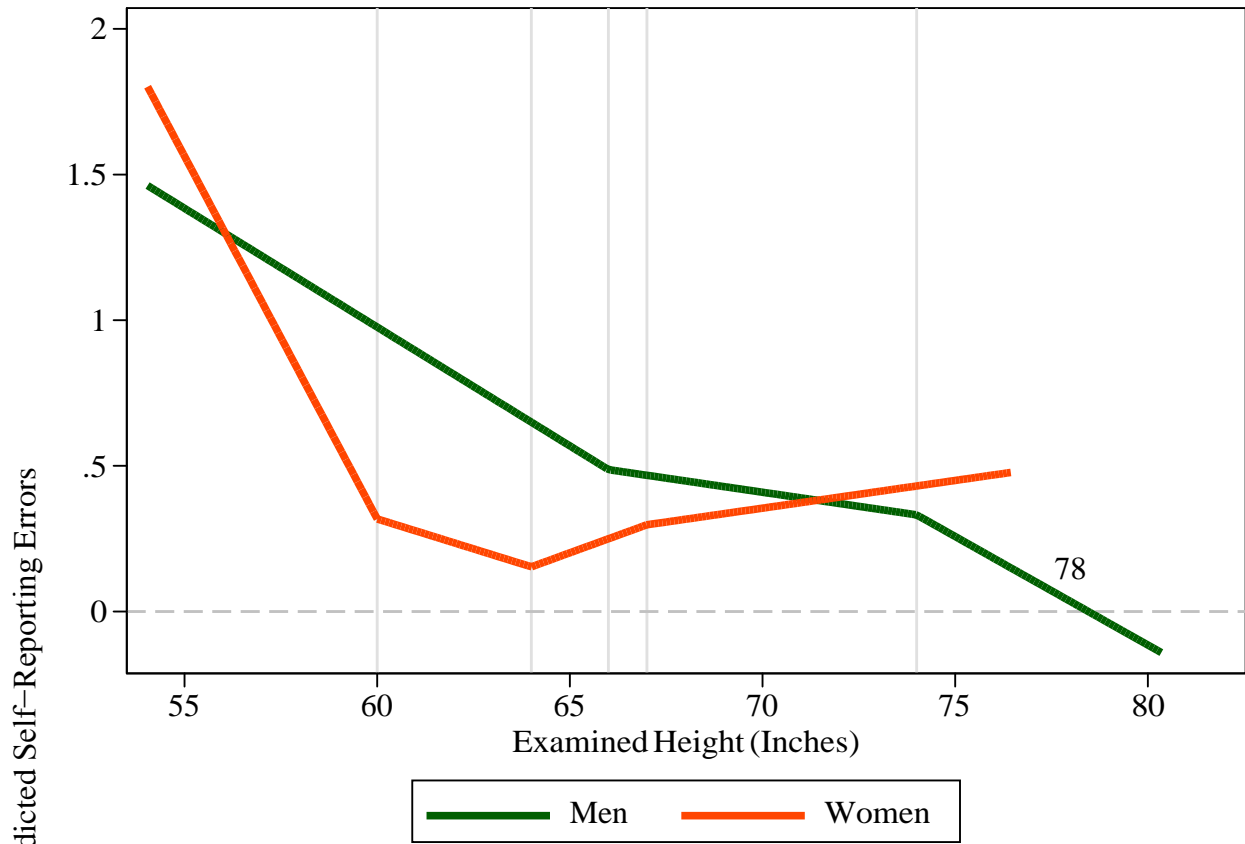
Source: Authors' calculations based on NHANES 1999–2010 data.

Figure 3: Fitted Values of Self-Reporting Errors in Weight



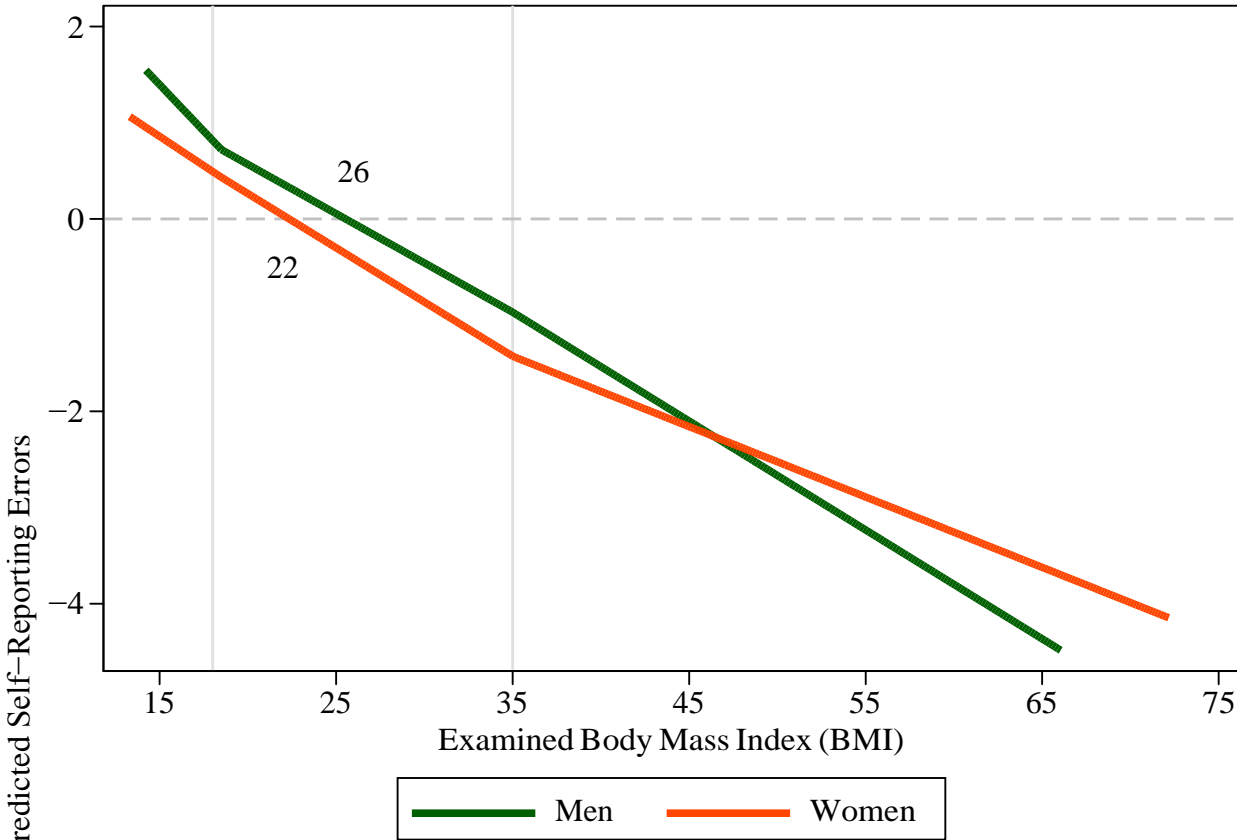
Source: Authors' calculations based on NHANES 1999–2010 data.

Figure 4: Fitted Values of Self-Reporting Errors in Height



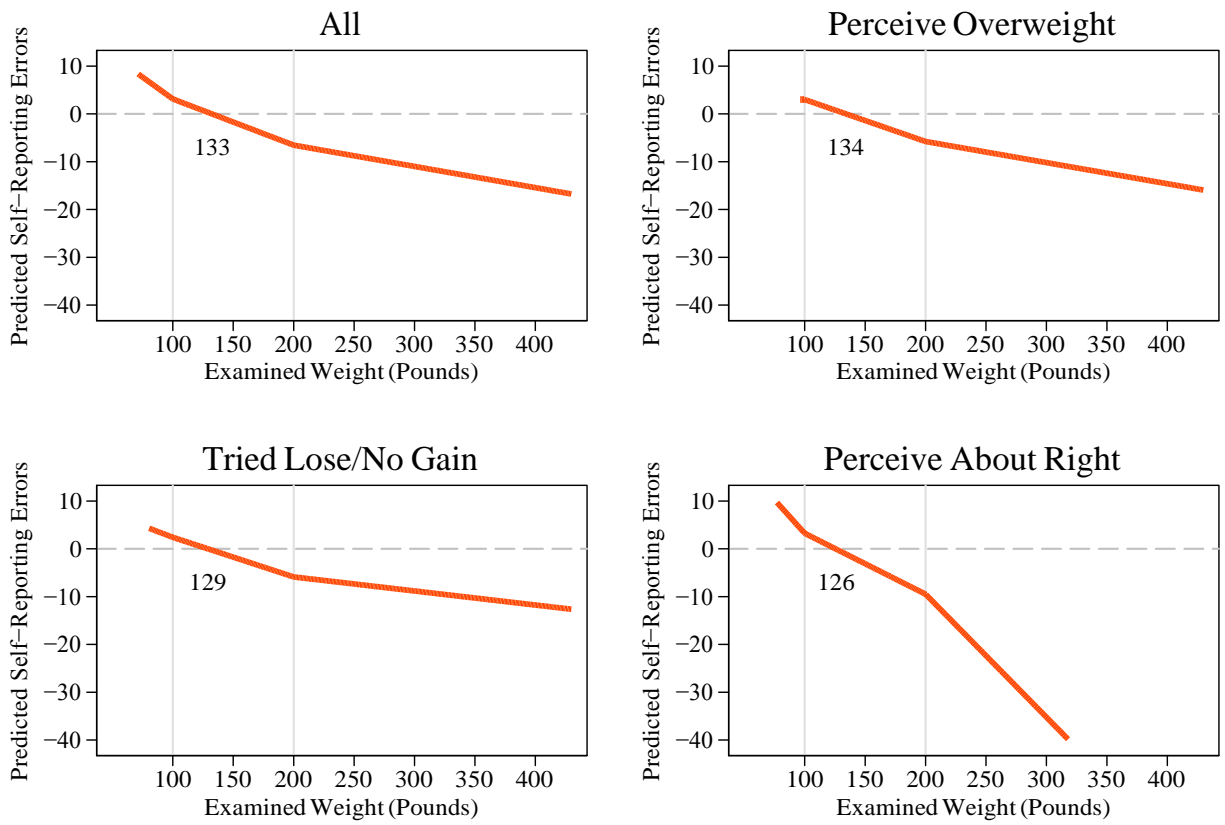
Source: Authors' calculations based on NHANES 1999–2010 data.

Figure 5: Fitted Values of Self-Reporting Errors in BMI



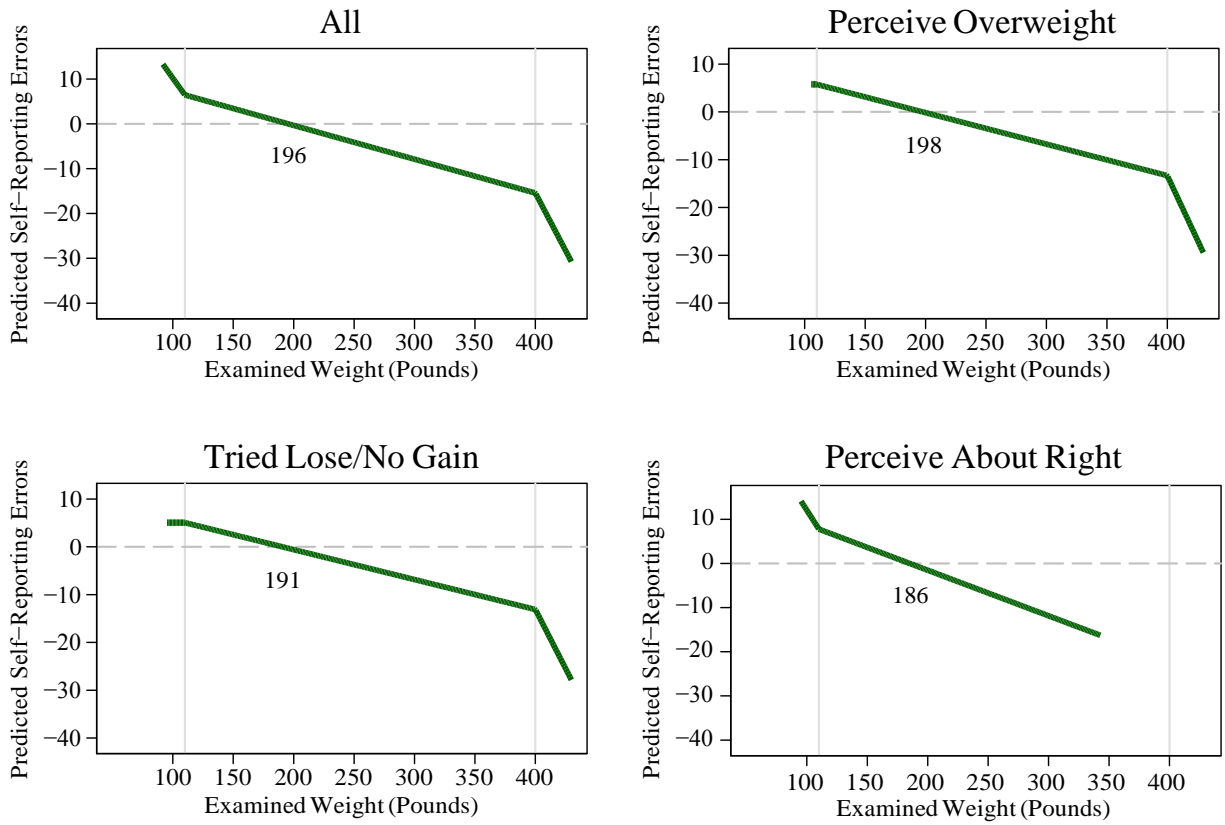
Source: Authors' calculations based on NHANES 1999–2010 data.

Figure 6: Fitted Values of Women's Self-Reporting Errors in Weight



Source: Authors' calculations based on NHANES 1999–2010 data.

Figure 7: Fitted Values of Men's Self-Reporting Errors in Weight



Source: Authors' calculations based on NHANES 1999–2010 data

Table 1: Summary Statistics

	Mean	SD	Women Median	Min	Max	Mean	SD	Men Median	Min	Max
X Weight(lb)	165.61	34.02	159.17	71.43	443.13	195.37	34.47	185.85	91.49	481.93
X Weight Squared	29,292.06	13,187.07	25,336.22	5,102.20	196,362.97	39,984.98	15,372.84	34,540.02	8,370.74	232,256.45
SR Weight(lb)	162.59	32.41	157.00	76.00	450.00	195.40	32.85	185.00	94.00	450.00
Self-Reported Errors, Weight(lb)	-3.02	6.58	-1.31	-72.84	48.32	0.04	6.92	0.84	-71.93	72.35
X Weight Neutral Rounding Error (5lbs)	-0.01	1.14	-0.00	-2.50	2.50	0.04	1.17	0.05	-2.50	2.49
X Height(in)	63.97	2.09	63.50	48.54	76.46	69.43	2.38	68.86	51.34	80.35
X Height(in) Squared	4,099.32	267.61	4,032.75	2,356.46	5,845.63	4,829.80	330.58	4,741.47	2,635.65	6,456.83
SR Height(in)	64.25	2.17	64.00	51.00	76.00	69.93	2.48	69.00	52.00	81.00
Self-Reported Errors, Height(in)	0.28	0.78	0.29	-3.98	5.50	0.49	0.82	0.55	-4.98	4.84
X Height Neutral Rounding Error (1in)	-0.00	0.23	-0.02	-0.50	0.50	-0.00	0.23	-0.00	-0.50	0.50
X BMI	28.44	5.61	27.94	13.18	72.28	28.43	4.57	27.72	14.20	66.16
X BMI Squared	859.57	367.96	780.64	173.71	5,224.40	840.17	296.24	768.40	201.64	4,377.15
SR BMI	27.68	5.26	27.12	13.46	71.82	28.04	4.25	27.32	14.25	70.48
Self-Reported Errors, BMI	-0.76	1.35	-0.57	-16.02	7.88	-0.39	1.22	-0.34	-12.62	10.92
Perceive Self Overweight	0.66	0.37	1.00	0.00	1.00	0.51	0.40	0.00	0.00	1.00
Perceive Self Underweight	0.03	0.13	0.00	0.00	1.00	0.06	0.19	0.00	0.00	1.00
Perceive Self About Right Weight	0.31	0.36	0.00	0.00	1.00	0.43	0.40	0.00	0.00	1.00
Time Lapse to Exam (Months)	0.84	0.47	1.00	0.00	4.00	0.83	0.48	1.00	0.00	4.00
Former Smoker	0.21	0.32	0.00	0.00	1.00	0.29	0.37	0.00	0.00	1.00
Current Smoker	0.22	0.32	0.00	0.00	1.00	0.27	0.36	0.00	0.00	1.00
20-34 Years	0.26	0.34	0.00	0.00	1.00	0.28	0.36	0.00	0.00	1.00
35-44 Years	0.22	0.33	0.00	0.00	1.00	0.22	0.34	0.00	0.00	1.00
45-59 Years	0.30	0.36	0.00	0.00	1.00	0.30	0.37	0.00	0.00	1.00
60-74 Years	0.17	0.30	0.00	0.00	1.00	0.15	0.29	0.00	0.00	1.00
75+ Years	0.05	0.18	0.00	0.00	1.00	0.04	0.17	0.00	0.00	1.00
All Hispanic	0.11	0.25	0.00	0.00	1.00	0.12	0.27	0.00	0.00	1.00
Non-Hispanic White	0.72	0.35	1.00	0.00	1.00	0.72	0.36	1.00	0.00	1.00
Non-Hispanic Black	0.12	0.25	0.00	0.00	1.00	0.10	0.24	0.00	0.00	1.00
Other/Multiracial	0.05	0.17	0.00	0.00	1.00	0.05	0.18	0.00	0.00	1.00
Never Married	0.14	0.28	0.00	0.00	1.00	0.17	0.31	0.00	0.00	1.00
Married/Partner	0.63	0.38	1.00	0.00	1.00	0.70	0.37	1.00	0.00	1.00
Widowed/Div/Sep	0.23	0.33	0.00	0.00	1.00	0.12	0.27	0.00	0.00	1.00
No High School Diploma	0.17	0.30	0.00	0.00	1.00	0.18	0.31	0.00	0.00	1.00
High School Diploma	0.25	0.34	0.00	0.00	1.00	0.26	0.35	0.00	0.00	1.00
Some College	0.32	0.37	0.00	0.00	1.00	0.28	0.37	0.00	0.00	1.00
College Degree or More	0.26	0.35	0.00	0.00	1.00	0.27	0.36	0.00	0.00	1.00
PIR* 0-1.85	0.32	0.37	0.00	0.00	1.00	0.27	0.36	0.00	0.00	1.00
PIR* 1.86-3.50	0.25	0.34	0.00	0.00	1.00	0.26	0.36	0.00	0.00	1.00
PIR* 3.51-4.99	0.18	0.30	0.00	0.00	1.00	0.19	0.32	0.00	0.00	1.00
PIR* 5+	0.24	0.34	0.00	0.00	1.00	0.28	0.36	0.00	0.00	1.00
Some Disability	0.20	0.32	0.00	0.00	1.00	0.18	0.31	0.00	0.00	1.00
Insured	0.84	0.29	1.00	0.00	1.00	0.80	0.32	1.00	0.00	1.00
1999 Survey	0.13	0.27	0.00	0.00	1.00	0.13	0.27	0.00	0.00	1.00
2001 Survey	0.17	0.30	0.00	0.00	1.00	0.17	0.30	0.00	0.00	1.00
2003 Survey	0.17	0.30	0.00	0.00	1.00	0.17	0.31	0.00	0.00	1.00
2005 Survey	0.18	0.30	0.00	0.00	1.00	0.18	0.31	0.00	0.00	1.00
2007 Survey	0.17	0.30	0.00	0.00	1.00	0.17	0.31	0.00	0.00	1.00
2009 Survey	0.18	0.30	0.00	0.00	1.00	0.18	0.31	0.00	0.00	1.00
Sample Size			11,507					11,819		

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.

Table 2: Weight Regressions: Women's Baseline Sample

	(1)	(2)	(3)	(4)
X Weight <100	-0.1252** (0.0601)	-0.1553** (0.0652)	-0.1827*** (0.0675)	-0.1895** (0.0892)
X Weight <200	-0.0733*** (0.0033)	-0.0808*** (0.0038)	-0.0967*** (0.0047)	-0.0969*** (0.0049)
X Weight >200	-0.0461*** (0.0090)	-0.0464*** (0.0091)	-0.0445*** (0.0091)	-0.0448*** (0.0104)
X Weight Neutral Rounding Error (5lbs)		0.1584*** (0.0499)	0.1633*** (0.0497)	0.1634*** (0.0497)
X Height(in)		0.6901 (0.9350)	0.7642 (0.9429)	0.7613 (0.9431)
X Height(in) Squared		-0.0040 (0.0074)	-0.0042 (0.0074)	-0.0042 (0.0074)
Time Lapse to Exam (Months)		-0.1478 (0.1290)	-0.1450 (0.1264)	-0.1511 (0.1314)
Former Smoker		-0.1172 (0.1803)	-0.1724 (0.1800)	-0.1728 (0.1803)
Current Smoker		-0.5652** (0.2154)	-0.5841*** (0.2128)	-0.5836*** (0.2134)
20-34 Years		-0.7038** (0.2734)	-0.6627** (0.2731)	-0.6623** (0.2731)
35-44 Years		-0.2464 (0.2656)	-0.2428 (0.2635)	-0.2425 (0.2635)
60-74 Years		0.4346* (0.2273)	0.5020** (0.2276)	0.5022** (0.2275)
75+ Years		0.4928 (0.3087)	0.6832** (0.3033)	0.6846** (0.3011)
All Hispanic		1.4421*** (0.2876)	1.4604*** (0.2883)	1.4607*** (0.2876)
Non-Hispanic Black		0.7770** (0.2967)	0.9265*** (0.2984)	0.9277*** (0.3005)
Other/Multiracial		-0.3246 (0.4317)	-0.3052 (0.4293)	-0.3051 (0.4295)
Never Married		-0.0279 (0.2836)	0.0421 (0.2839)	0.0424 (0.2830)
Widowed/Div/Sep		0.3908* (0.2176)	0.4338** (0.2171)	0.4346** (0.2172)
No High School Diploma		0.0877 (0.3032)	0.1806 (0.3053)	0.1797 (0.3050)
Some College		-0.4176* (0.2335)	-0.4372* (0.2356)	-0.4376* (0.2360)
College Degree or More		-0.2801 (0.2924)	-0.2901 (0.2936)	-0.2904 (0.2939)
PIR* 1.86-3.50		-0.0490 (0.2827)	-0.1229 (0.2802)	-0.1232 (0.2814)
PIR* 3.51-4.99		0.2079 (0.2348)	0.0996 (0.2295)	0.0992 (0.2297)
PIR* 5+		0.0074 (0.2675)	-0.0753 (0.2640)	-0.0758 (0.2642)
Some Disability		0.4991** (0.2172)	0.4711** (0.2184)	0.4711** (0.2186)
Insured		0.7461** (0.3217)	0.7566** (0.3245)	0.7566** (0.3245)
2001 Survey		0.3708 (0.3561)	0.3878 (0.3544)	0.3875 (0.3540)
2003 Survey		0.9125** (0.3800)	0.9286** (0.3844)	0.9285** (0.3843)
2005 Survey		0.6582** (0.3054)	0.6905** (0.3063)	0.6899** (0.3054)
2007 Survey		0.6233* (0.3394)	0.6920** (0.3435)	0.6919** (0.3434)
2009 Survey		0.9522*** (0.3084)	1.0278*** (0.3071)	1.0273*** (0.3069)
Perceive Self Underweight			-1.7447*** (0.3727)	-1.7502*** (0.3776)
Perceive Self About Right Weight			-1.3939*** (0.2195)	-1.3964*** (0.2244)
[X Weight <100] x Time Lapse				-0.0908 (0.6054)
[X Weight >200] x Time Lapse				0.0374 (0.4178)
R ²	0.11	0.13	0.13	0.13
Sample Size	11,507	11,507	11,507	11,507

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.

Table 3: Weight Regressions: Men's Baseline Sample

	(1)	(2)	(3)	(4)
X Weight <110	-0.3546** (0.1683)	-0.3958*** (0.1463)	-0.3768** (0.1461)	-0.2975* (0.1746)
X Weight <400	-0.0583*** (0.0034)	-0.0700*** (0.0046)	-0.0755*** (0.0059)	-0.0755*** (0.0060)
X Weight >400	-0.5410*** (0.1081)	-0.5254*** (0.1053)	-0.5131*** (0.1053)	-0.5134*** (0.1569)
X Weight Neutral Rounding Error (5lbs)		0.0785 (0.0702)	0.0725 (0.0701)	0.0724 (0.0703)
X Height(in)		-2.1091* (1.1765)	-2.2093* (1.1690)	-2.1871* (1.1750)
X Height(in) Squared		0.0180** (0.0084)	0.0190** (0.0084)	0.0188** (0.0084)
Time Lapse to Exam (Months)		-0.2749** (0.1370)	-0.2738** (0.1378)	-0.2757** (0.1361)
Former Smoker		0.1556 (0.2157)	0.1293 (0.2156)	0.1294 (0.2158)
Current Smoker		0.9415*** (0.2402)	0.9244*** (0.2394)	0.9242*** (0.2390)
20-34 Years		-0.2779 (0.2404)	-0.2336 (0.2416)	-0.2332 (0.2419)
35-44 Years		0.1273 (0.2249)	0.1422 (0.2286)	0.1425 (0.2273)
60-74 Years		0.1391 (0.2363)	0.1488 (0.2352)	0.1495 (0.2352)
75+ Years		0.0121 (0.3447)	0.0804 (0.3505)	0.0818 (0.3503)
All Hispanic		-1.0334*** (0.3196)	-1.0050*** (0.3240)	-1.0024*** (0.3230)
Non-Hispanic Black		0.4870* (0.2749)	0.5880** (0.2741)	0.5881** (0.2739)
Other/Multiracial		-1.0045** (0.4522)	-0.9809** (0.4532)	-0.9772** (0.4545)
Never Married		-0.8840*** (0.2869)	-0.8734*** (0.2868)	-0.8752*** (0.2862)
Widowed/Div/Sep		0.0865 (0.3360)	0.1091 (0.3347)	0.1100 (0.3347)
No High School Diploma		0.5211* (0.2625)	0.5620** (0.2634)	0.5618** (0.2618)
Some College		-0.4912** (0.2129)	-0.4997** (0.2143)	-0.5002** (0.2141)
College Degree or More		-1.5913*** (0.2584)	-1.6118*** (0.2643)	-1.6116*** (0.2637)
PIR* 1.86-3.50		-0.1547 (0.2573)	-0.1716 (0.2552)	-0.1727 (0.2547)
PIR* 3.51-4.99		0.3294 (0.3082)	0.3027 (0.3039)	0.3022 (0.3044)
PIR* 5+		-0.4562* (0.2585)	-0.4951* (0.2525)	-0.4953* (0.2519)
Some Disability		0.4396** (0.2206)	0.4304* (0.2214)	0.4299* (0.2214)
Insured		-0.0663 (0.2993)	-0.0761 (0.2963)	-0.0757 (0.2962)
2001 Survey		0.1577 (0.3012)	0.1603 (0.3022)	0.1589 (0.3014)
2003 Survey		0.3201 (0.3350)	0.3174 (0.3374)	0.3174 (0.3369)
2005 Survey		0.1107 (0.3290)	0.1130 (0.3356)	0.1122 (0.3357)
2007 Survey		0.2078 (0.3049)	0.2187 (0.3087)	0.2164 (0.3086)
2009 Survey		0.0575 (0.3133)	0.0827 (0.3136)	0.0799 (0.3144)
Perceive Self Underweight			-0.3666 (0.5401)	-0.3749 (0.5435)
Perceive Self About Right Weight			-0.8189*** (0.2498)	-0.8187*** (0.2521)
[X Weight <110] x Time Lapse				1.0079 (1.6605)
[X Weight >400] x Time Lapse				0.0175 (5.6530)
R ²	0.11	0.15	0.15	0.15
Sample Size	11,819	11,819	11,819	11,819

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.

Table 4: Height Regressions, Baseline Sample

	Women		Men	
	(1)	(2)	(1)	(2)
X Height <60	-0.2772*** (0.0361)	-0.2490*** (0.0335)		
X Height <64	-0.0709*** (0.0115)	-0.0411*** (0.0108)		
X Height <67	0.0429*** (0.0127)	0.0481*** (0.0118)		
X Height >67	0.0108 (0.0173)	0.0190 (0.0173)		
X Height <66			-0.0617*** (0.0187)	-0.0816*** (0.0206)
X Height <74			-0.0117** (0.0048)	-0.0194*** (0.0051)
X Height >74			-0.0857*** (0.0192)	-0.0745*** (0.0188)
X Height Neutral Rounding Error (1in)		0.0983*** (0.0297)		0.0484 (0.0325)
X Weight(lb)		0.0032** (0.0014)		0.0022 (0.0016)
X Weight Squared		-0.0000 (0.0000)		0.0000 (0.0000)
Time Lapse to Exam(months)		0.0042 (0.0154)		0.0245 (0.0180)
Former Smoker		-0.0404* (0.0231)		0.0100 (0.0244)
Current Smoker		-0.1025*** (0.0277)		0.1302*** (0.0252)
20-34 Years		-0.0241 (0.0244)		-0.0108 (0.0263)
35-44 Years		-0.0806*** (0.0251)		-0.0664** (0.0267)
60-74 Years		0.2780*** (0.0244)		0.3507*** (0.0289)
75+ Years		0.5695*** (0.0578)		0.7253*** (0.0481)
All Hispanic		0.2182*** (0.0394)		-0.2066*** (0.0373)
Non-Hispanic Black		-0.0122 (0.0237)		-0.0242 (0.0314)
Other/Multiracial		0.1081** (0.0427)		-0.1446*** (0.0506)
Never Married		0.0749*** (0.0272)		0.0178 (0.0307)
Widowed/Div/Sep		0.0331 (0.0251)		0.0883*** (0.0326)
No High School Diploma		0.1047*** (0.0353)		-0.0404 (0.0384)
Some College		0.0025 (0.0247)		0.0283 (0.0289)
College Degree or More		-0.0210 (0.0298)		-0.0760** (0.0333)
PIR* 1.86-3.50		-0.0469 (0.0322)		0.0321 (0.0346)
PIR* 3.51-4.99		-0.0776** (0.0337)		0.0348 (0.0360)
PIR* 5+		-0.1332*** (0.0367)		-0.0417 (0.0320)
Some Disability		0.0121 (0.0220)		0.0472 (0.0296)
Insured		-0.0176 (0.0336)		-0.0438 (0.0344)
2001 Survey		-0.0719* (0.0379)		-0.0483 (0.0447)
2003 Survey		-0.0155 (0.0421)		0.0035 (0.0479)
2005 Survey		0.0053 (0.0369)		-0.0044 (0.0428)
2007 Survey		-0.0091 (0.0365)		-0.0023 (0.0448)
2009 Survey		0.0276 (0.0392)		-0.0009 (0.0436)
R ²	0.03	0.10	0.01	0.07
Sample Size	11,507	11,507	11,819	11,819

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.

Table 5: BMI Regressions: Women's Baseline Sample

	(1)	(2)	(3)	(4)
X BMI < 18.5	-0.0476 (0.0447)	-0.0797* (0.0440)	-0.1216** (0.0484)	-0.1482** (0.0597)
X BMI < 35	-0.0896*** (0.0040)	-0.0912*** (0.0043)	-0.1127*** (0.0055)	-0.1130*** (0.0057)
X BMI > 35	-0.0761*** (0.0111)	-0.0778*** (0.0112)	-0.0732*** (0.0113)	-0.0733*** (0.0137)
Time Lapse to Exam (Months)		-0.0308 (0.0270)	-0.0301 (0.0267)	-0.0295 (0.0248)
Former Smoker		0.0342 (0.0346)	0.0200 (0.0347)	0.0198 (0.0350)
Current Smoker		-0.0210 (0.0471)	-0.0268 (0.0467)	-0.0271 (0.0465)
20-34 Years		-0.0787 (0.0523)	-0.0685 (0.0521)	-0.0687 (0.0520)
35-44 Years		0.0378 (0.0492)	0.0386 (0.0489)	0.0385 (0.0488)
60-74 Years		-0.1381*** (0.0441)	-0.1210*** (0.0449)	-0.1212*** (0.0448)
75+ Years		-0.3373*** (0.0646)	-0.2867*** (0.0659)	-0.2873*** (0.0661)
All Hispanic		0.0520 (0.0646)	0.0608 (0.0644)	0.0607 (0.0643)
Non-Hispanic Black		0.1126** (0.0536)	0.1485*** (0.0540)	0.1487*** (0.0544)
Other/Multiracial		-0.0618 (0.0848)	-0.0505 (0.0840)	-0.0510 (0.0837)
Never Married		-0.0765 (0.0518)	-0.0598 (0.0516)	-0.0594 (0.0516)
Widowed/Div/Sep		0.0392 (0.0406)	0.0496 (0.0405)	0.0500 (0.0405)
No High School Diploma		-0.0762 (0.0618)	-0.0532 (0.0617)	-0.0534 (0.0618)
Some College		-0.0626 (0.0439)	-0.0689 (0.0443)	-0.0688 (0.0443)
College Degree or More		-0.0790 (0.0560)	-0.0828 (0.0557)	-0.0825 (0.0556)
PIR* 1.86-3.50		0.0390 (0.0529)	0.0205 (0.0526)	0.0207 (0.0528)
PIR* 3.51-4.99		0.0844* (0.0505)	0.0573 (0.0500)	0.0574 (0.0498)
PIR* 5+		0.1175** (0.0544)	0.0964* (0.0540)	0.0963* (0.0538)
Some Disability		0.0566 (0.0420)	0.0489 (0.0419)	0.0488 (0.0420)
Insured		0.1463** (0.0629)	0.1493** (0.0627)	0.1493** (0.0627)
2001 Survey		0.0521 (0.0503)	0.0560 (0.0488)	0.0560 (0.0490)
2003 Survey		0.1089* (0.0614)	0.1129* (0.0606)	0.1124* (0.0606)
2005 Survey		0.0479 (0.0545)	0.0564 (0.0539)	0.0559 (0.0541)
2007 Survey		0.0687 (0.0554)	0.0857 (0.0554)	0.0851 (0.0556)
2009 Survey		0.0646 (0.0440)	0.0827* (0.0429)	0.0826* (0.0432)
Perceive Self Underweight			-0.4223*** (0.0738)	-0.4192*** (0.0743)
Perceive Self About Right Weight			-0.3390*** (0.0442)	-0.3388*** (0.0442)
[X BMI < 18.5] x Time Lapse				-0.0639 (0.0802)
[X BMI > 35] x Time Lapse				0.0043 (0.0940)
R ²	0.13	0.14	0.15	0.15
Sample Size	11,507	11,507	11,507	11,507

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.

Table 6: BMI Regressions: Men's Baseline Sample

	(1)	(2)	(3)	(4)
X BMI < 18.5	-0.2203** (0.1009)	-0.2160** (0.0996)	-0.1930* (0.0986)	-0.2139** (0.0981)
X BMI < 35	-0.0802*** (0.0042)	-0.0824*** (0.0043)	-0.1024*** (0.0052)	-0.0984*** (0.0055)
X BMI > 35	-0.1126*** (0.0228)	-0.1150*** (0.0227)	-0.1132*** (0.0227)	-0.1039*** (0.0250)
Time Lapse to Exam (Months)		-0.0456* (0.0242)	-0.0452* (0.0244)	-0.0288 (0.0224)
Former Smoker		-0.0150 (0.0347)	-0.0204 (0.0341)	-0.0223 (0.0338)
Current Smoker		-0.0308 (0.0420)	-0.0319 (0.0422)	-0.0310 (0.0421)
20-34 Years		-0.0663* (0.0377)	-0.0536 (0.0375)	-0.0543 (0.0374)
35-44 Years		0.0492 (0.0341)	0.0566 (0.0345)	0.0557 (0.0345)
60-74 Years		-0.2297*** (0.0381)	-0.2253*** (0.0377)	-0.2243*** (0.0373)
75+ Years		-0.4724*** (0.0621)	-0.4508*** (0.0617)	-0.4517*** (0.0615)
All Hispanic		0.0080 (0.0447)	0.0183 (0.0449)	0.0173 (0.0448)
Non-Hispanic Black		0.0709 (0.0455)	0.1040** (0.0458)	0.1051** (0.0457)
Other/Multiracial		-0.0437 (0.0865)	-0.0410 (0.0868)	-0.0415 (0.0869)
Never Married		-0.1162*** (0.0388)	-0.1128*** (0.0386)	-0.1115*** (0.0387)
Widowed/Div/Sep		-0.0475 (0.0469)	-0.0420 (0.0466)	-0.0443 (0.0466)
No High School Diploma		0.0604 (0.0467)	0.0714 (0.0470)	0.0720 (0.0469)
Some College		-0.1096*** (0.0342)	-0.1118*** (0.0346)	-0.1104*** (0.0346)
College Degree or More		-0.1277*** (0.0436)	-0.1408*** (0.0438)	-0.1390*** (0.0440)
PIR* 1.86-3.50		-0.0318 (0.0440)	-0.0382 (0.0442)	-0.0402 (0.0445)
PIR* 3.51-4.99		0.0158 (0.0438)	0.0050 (0.0434)	0.0041 (0.0436)
PIR* 5+		-0.0132 (0.0409)	-0.0289 (0.0411)	-0.0328 (0.0417)
Some Disability		0.0340 (0.0393)	0.0339 (0.0390)	0.0342 (0.0387)
Insured		0.0238 (0.0469)	0.0216 (0.0468)	0.0223 (0.0467)
2001 Survey		0.0437 (0.0485)	0.0423 (0.0480)	0.0385 (0.0486)
2003 Survey		0.0286 (0.0468)	0.0292 (0.0480)	0.0254 (0.0483)
2005 Survey		0.0550 (0.0521)	0.0546 (0.0539)	0.0517 (0.0546)
2007 Survey		0.0219 (0.0499)	0.0257 (0.0501)	0.0231 (0.0504)
2009 Survey		0.0116 (0.0471)	0.0186 (0.0475)	0.0159 (0.0481)
Perceive Self Underweight			-0.3082*** (0.0595)	-0.2888*** (0.0614)
Perceive Self About Right Weight			-0.2577*** (0.0368)	-0.2518*** (0.0375)
[X BMI < 18.5] x Time Lapse				-0.0283 (0.1109)
[X BMI > 35] x Time Lapse				-0.1562* (0.0922)
R ²	0.13	0.14	0.15	0.15
Sample Size	11,819	11,819	11,819	11,819

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.

Table 7: Weight Regressions: Women's Baseline Sample

	Perceive Overweight	Perceive Overweight	Tried Lose/No Gain	Tried Lose/No Gain	Perceive About Right	Perceive About Right
X Weight <100			-0.0974*	0.0292	-0.2812***	-0.2794**
			(0.0572)	(0.0964)	(0.0905)	(0.1111)
X Weight <200	-0.0877***	-0.0857***	-0.0829***	-0.0802***	-0.1273***	-0.1280***
	(0.0055)	(0.0056)	(0.0055)	(0.0054)	(0.0093)	(0.0096)
X Weight >200	-0.0442***	-0.0424***	-0.0294***	-0.0263**	-0.2575***	-0.2681***
	(0.0099)	(0.0110)	(0.0106)	(0.0124)	(0.0659)	(0.0673)
X Weight Neutral Rounding Error (5lbs)	0.2127***	0.2124***	0.2140***	0.2132***	0.0580	0.0565
	(0.0658)	(0.0659)	(0.0757)	(0.0759)	(0.0686)	(0.0688)
X Height(in)	1.6913	1.6770	0.4302	0.4233	-1.8422	-1.8541
	(1.2738)	(1.2721)	(1.2282)	(1.2230)	(1.4040)	(1.4096)
X Height(in) Squared	-0.0119	-0.0118	-0.0019	-0.0018	0.0177	0.0178
	(0.0100)	(0.0100)	(0.0097)	(0.0097)	(0.0111)	(0.0112)
Time Lapse to Exam (Months)	-0.0572	0.0186	-0.1725	-0.0759	-0.3046*	-0.3131*
	(0.1837)	(0.1919)	(0.1878)	(0.1926)	(0.1677)	(0.1729)
Former Smoker	-0.3767	-0.3721	-0.1426	-0.1396	0.2145	0.2104
	(0.2403)	(0.2405)	(0.2602)	(0.2629)	(0.2527)	(0.2549)
Current Smoker	-0.8764***	-0.8709***	-0.7270**	-0.7242**	0.1448	0.1478
	(0.3215)	(0.3197)	(0.3630)	(0.3634)	(0.2889)	(0.2906)
20-34 Years	-0.8115**	-0.8132**	-1.1888***	-1.1917***	-0.3576	-0.3598
	(0.3942)	(0.3940)	(0.3487)	(0.3501)	(0.3760)	(0.3767)
35-44 Years	-0.1029	-0.1022	-0.3378	-0.3358	-0.3527	-0.3540
	(0.3378)	(0.3373)	(0.3257)	(0.3254)	(0.3553)	(0.3560)
60-74 Years	0.4458	0.4430	0.2085	0.2032	1.0053**	1.0003**
	(0.3058)	(0.3056)	(0.3094)	(0.3096)	(0.3898)	(0.3873)
75+ Years	0.9527**	0.9472**	0.6510	0.6377	1.1629**	1.1658**
	(0.3991)	(0.3979)	(0.4123)	(0.4101)	(0.4788)	(0.4790)
All Hispanic	1.6912***	1.6898***	2.0416***	2.0360***	1.1824***	1.1783***
	(0.3903)	(0.3909)	(0.3058)	(0.3060)	(0.3350)	(0.3329)
Non-Hispanic Black	0.4598	0.4533	0.6039	0.5908	2.3031***	2.3050***
	(0.3762)	(0.3776)	(0.3995)	(0.3987)	(0.4507)	(0.4482)
Other/Multiracial	-0.2626	-0.2621	-0.5864	-0.5962	-0.2505	-0.2532
	(0.7147)	(0.7147)	(0.7278)	(0.7307)	(0.4134)	(0.4134)
Never Married	0.3220	0.3131	0.1240	0.1180	-0.5978**	-0.5972**
	(0.4240)	(0.4216)	(0.4123)	(0.4081)	(0.2758)	(0.2755)
Widowed/Div/Sep	0.8497***	0.8471***	0.6077*	0.6061*	-0.5930**	-0.5978**
	(0.2853)	(0.2854)	(0.3137)	(0.3139)	(0.2942)	(0.2959)
No High School Diploma	0.1898	0.1958	0.0636	0.0685	0.5788	0.5833
	(0.4417)	(0.4410)	(0.5124)	(0.5107)	(0.3576)	(0.3587)
Some College	-0.5682*	-0.5628*	-0.7223**	-0.7196**	-0.2154	-0.2129
	(0.3022)	(0.3027)	(0.2939)	(0.2943)	(0.2965)	(0.2953)
College Degree or More	-0.2273	-0.2251	-0.6583*	-0.6585*	-0.4344	-0.4335
	(0.4033)	(0.4032)	(0.3631)	(0.3628)	(0.2954)	(0.2946)
PIR* 1.86-3.50	-0.0718	-0.0679	0.0820	0.0837	-0.2438	-0.2430
	(0.3858)	(0.3871)	(0.3771)	(0.3777)	(0.3307)	(0.3308)
PIR* 3.51-4.99	0.1116	0.1159	0.7088**	0.7085**	0.0476	0.0479
	(0.3184)	(0.3185)	(0.3215)	(0.3219)	(0.3660)	(0.3656)
PIR* 5+	0.1467	0.1500	0.4306	0.4322	-0.5740*	-0.5742*
	(0.3775)	(0.3783)	(0.3322)	(0.3331)	(0.3329)	(0.3323)
Some Disability	0.7230**	0.7235**	0.8342***	0.8335***	-0.2739	-0.2761
	(0.2763)	(0.2761)	(0.2844)	(0.2845)	(0.3512)	(0.3508)
Insured	1.1383**	1.1389**	0.6736	0.6711	0.2809	0.2782
	(0.4912)	(0.4906)	(0.4795)	(0.4783)	(0.3286)	(0.3276)
2001 Survey	0.5012	0.5084	0.6118	0.6158	-0.1339	-0.1358
	(0.4352)	(0.4354)	(0.3858)	(0.3863)	(0.4782)	(0.4781)
2003 Survey	1.0548**	1.0555**	1.5864***	1.5867***	0.6682	0.6681
	(0.4941)	(0.4941)	(0.4952)	(0.4944)	(0.4804)	(0.4798)
2005 Survey	0.9167**	0.9213**	1.0374**	1.0441**	0.1215	0.1216
	(0.3918)	(0.3912)	(0.4112)	(0.4122)	(0.4238)	(0.4257)
2007 Survey	0.6634	0.6666	0.8522*	0.8512*	0.6176	0.6188
	(0.4106)	(0.4110)	(0.4358)	(0.4377)	(0.4574)	(0.4576)
2009 Survey	1.2149***	1.2252***	1.4731***	1.4828***	0.5941	0.5960
	(0.3942)	(0.3932)	(0.4169)	(0.4174)	(0.4462)	(0.4447)
[X Weight <100] x Time Lapse				2.0614		-0.0005
				(1.6565)		(0.6476)
[X Weight >200] x Time Lapse		-0.2659		-0.4287		1.0414
		(0.4196)		(0.5363)		(2.4554)
R ²	0.10	0.10	0.11	0.11	0.20	0.20
Sample Size	7,526	7,526	5,858	5,858	3,593	3,593

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.

Table 8: Weight Regressions: Men's Baseline Sample

	Perceive Overweight	Perceive Overweight	Tried Lose/No Gain	Tried Lose/No Gain	Perceive About Right	Perceive About Right
X Weight <110					-0.4292*	0.0542
					(0.2212)	(0.3534)
X Weight <400	-0.0658***	-0.0658***	-0.0626***	-0.0620***	-0.1033***	-0.1031***
	(0.0081)	(0.0082)	(0.0056)	(0.0055)	(0.0062)	(0.0061)
X Weight >400	-0.5400***	-0.5340***	-0.4940***	-0.2174*		
	(0.1068)	(0.1543)	(0.1079)	(0.1104)		
X Weight Neutral Rounding Error (5lbs)	0.1423	0.1422	0.1319	0.1325	0.0724	0.0721
	(0.0994)	(0.0996)	(0.0950)	(0.0950)	(0.0900)	(0.0900)
X Height(in)	-1.9386	-1.9386	-1.4607	-1.4686	-1.5472	-1.3336
	(1.6608)	(1.6608)	(1.6388)	(1.6244)	(1.2216)	(1.2387)
X Height(in) Squared	0.0167	0.0166	0.0132	0.0132	0.0152*	0.0137
	(0.0118)	(0.0118)	(0.0118)	(0.0117)	(0.0087)	(0.0089)
Time Lapse to Exam (Months)	-0.4484**	-0.4471**	-0.4967**	-0.4555**	0.0389	0.0321
	(0.2043)	(0.1992)	(0.2054)	(0.1998)	(0.2048)	(0.2056)
Former Smoker	0.1734	0.1746	0.0424	0.0416	0.2355	0.2308
	(0.3139)	(0.3147)	(0.2564)	(0.2547)	(0.2694)	(0.2679)
Current Smoker	0.5175	0.5185	0.8545**	0.8793**	1.2597***	1.2616***
	(0.4016)	(0.4025)	(0.3930)	(0.3930)	(0.3101)	(0.3100)
20-34 Years	-0.6399*	-0.6396*	-0.5365	-0.5083	0.2195	0.2200
	(0.3771)	(0.3770)	(0.3933)	(0.3838)	(0.3451)	(0.3446)
35-44 Years	0.1762	0.1756	-0.1333	-0.1443	0.2461	0.2490
	(0.3414)	(0.3400)	(0.3472)	(0.3420)	(0.3511)	(0.3511)
60-74 Years	-0.2848	-0.2850	-0.0929	-0.0491	0.7717**	0.7752**
	(0.3796)	(0.3802)	(0.3347)	(0.3320)	(0.3249)	(0.3270)
75+ Years	-0.5435	-0.5436	-0.1434	-0.1196	1.1248**	1.1237**
	(0.5659)	(0.5663)	(0.4620)	(0.4567)	(0.4384)	(0.4368)
All Hispanic	-1.2924***	-1.2917***	-1.1644***	-1.1558***	-0.5883	-0.5757
	(0.4810)	(0.4803)	(0.4398)	(0.4372)	(0.4016)	(0.3988)
Non-Hispanic Black	-0.4992	-0.4983	0.2734	0.2888	1.2329***	1.2320***
	(0.4107)	(0.4111)	(0.4123)	(0.4105)	(0.3887)	(0.3885)
Other/Multiracial	-1.2506*	-1.2506*	-1.5837*	-1.5942*	-0.5788	-0.5546
	(0.7136)	(0.7136)	(0.8440)	(0.8440)	(0.5954)	(0.5975)
Never Married	-2.2360***	-2.2353***	-1.2254***	-1.2052***	0.2131	0.2156
	(0.4809)	(0.4804)	(0.4523)	(0.4478)	(0.3201)	(0.3204)
Widowed/Div/Sep	-0.0501	-0.0508	0.0004	-0.0443	0.4347	0.4392
	(0.4723)	(0.4715)	(0.4805)	(0.4745)	(0.3673)	(0.3674)
No High School Diploma	0.6218	0.6195	0.5490	0.4307	1.1424***	1.1421***
	(0.3869)	(0.3807)	(0.4607)	(0.4499)	(0.4045)	(0.4045)
Some College	-0.9824***	-0.9836***	-0.3772	-0.3846	0.2142	0.2054
	(0.2994)	(0.2976)	(0.3645)	(0.3626)	(0.3633)	(0.3636)
College Degree or More	-1.5971***	-1.5977***	-1.6216***	-1.6304***	-1.6644***	-1.6659***
	(0.4337)	(0.4323)	(0.3596)	(0.3579)	(0.3603)	(0.3602)
PIR* 1.86-3.50	-0.5240	-0.5249	-0.7012	-0.7428*	0.3634	0.3606
	(0.4328)	(0.4323)	(0.4487)	(0.4429)	(0.3486)	(0.3490)
PIR* 3.51-4.99	0.0839	0.0831	0.1019	0.0329	0.8588*	0.8617*
	(0.4327)	(0.4337)	(0.4321)	(0.4316)	(0.4716)	(0.4712)
PIR* 5+	-0.7344	-0.7353	-0.5176	-0.5671	-0.0766	-0.0758
	(0.4470)	(0.4456)	(0.4263)	(0.4220)	(0.4206)	(0.4201)
Some Disability	0.8701***	0.8701***	0.7322**	0.7599**	-0.2623	-0.2628
	(0.3223)	(0.3223)	(0.3501)	(0.3484)	(0.3182)	(0.3166)
Insured	-0.0754	-0.0744	-0.0695	-0.0450	0.0376	0.0418
	(0.4547)	(0.4549)	(0.4654)	(0.4527)	(0.3297)	(0.3296)
2001 Survey	0.2146	0.2158	-0.3956	-0.3703	0.1311	0.1292
	(0.5006)	(0.4981)	(0.5215)	(0.5079)	(0.3451)	(0.3453)
2003 Survey	0.5049	0.5059	0.4781	0.5258	0.4562	0.4534
	(0.5100)	(0.5076)	(0.5315)	(0.5120)	(0.4096)	(0.4102)
2005 Survey	0.0635	0.0639	-0.0043	0.0294	0.1641	0.1649
	(0.3622)	(0.3613)	(0.5893)	(0.5751)	(0.4390)	(0.4382)
2007 Survey	0.2860	0.2868	0.1729	0.2132	0.4880	0.4794
	(0.4513)	(0.4501)	(0.5662)	(0.5457)	(0.4375)	(0.4386)
2009 Survey	0.5637	0.5648	0.3060	0.3547	-0.3281	-0.3454
	(0.4664)	(0.4643)	(0.5165)	(0.4990)	(0.4503)	(0.4476)
[X Weight <110] x Time Lapse						6.2832**
						(2.8818)
[X Weight >400] x Time Lapse		-0.3624		-14.8862***		
		(5.4904)		(5.0559)		
R ²	0.14	0.14	0.15	0.15	0.13	0.13
Sample Size	5,620	5,620	4,312	4,312	5,456	5,456

* PIR stands for Poverty Income Ratio. This gives the ratio of the family's household income to the poverty-income level for a household of the given size.