



Published in final edited form as:

Matern Child Health J. 2009 July ; 13(4): 489–496. doi:10.1007/s10995-008-0384-7.

Accuracy of Body Mass Index Categories Based on Self-Reported Height and Weight Among Women in the United States

Benjamin M. Craig and

Moffitt Cancer Center, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612-9416, USA

Alexandra K. Adams

Department of Family Medicine, University of Wisconsin-Madison, 777 South Mills Street, Room 3818, Madison, WI 53715, USA, e-mail: alex.adams@fammed.wisc.edu

Abstract

Objective—The purpose of this study was to assess the accuracy of BMI categories based on self-reported height and weight in adult women.

Methods—BMI categories from self-reported responses were compared to categories measured during physical examination from women, age 18 or older, who participated in the National Health and Examination Survey, 1999-2004. We first examined strength of agreement using Cohen's kappa, which, unlike sensitivity and specificity, allows for the comparison of polychotomous measures beyond chance agreement. Kappa regression identifies potential threats to accuracy. Likelihood of bias, as measured by under-reporting, was examined using logistic regression.

Results—Cohen's kappa estimates were 0.443 for pregnant women ($N = 724$) and 0.705 for non-pregnant women ($N = 5,910$). Kappa varied by age and race, but was largely unrelated to socioeconomic status, health and health behaviors. Women who visited a physician in the last year or been diagnosed with osteoporosis were more accurate, while women most likely to under-report were older, white, non-Hispanic, and college-educated.

Conclusions—Our results suggest substantial agreement between self-reported and measured categories, except for women who are pregnant, above the age of 75 or without physician visits. Under-reporting may be more prevalent in well-educated, white populations than minority populations.

Keywords

Obesity; Body mass index; Cohen's kappa

Obesity is a major public health epidemic and is an important risk factor contributing to morbidity and mortality from diseases such as heart disease, diabetes and cancer [1]. One of the challenges facing epidemiologists studying trends in the obesity epidemic is tracking changes over time. Both epidemiologists and clinicians often rely on self-reported height and weight, which are then used to calculate body mass index (BMI). Many studies have examined the accuracy of self-reported height, weight and BMI in a variety of cohorts [2-6]. Review of the literature indicates high correlations between self-reported and measured height and weight. However, studies suggest that accuracy may vary significantly according to age, gender, and socioeconomic status (SES) [2].

Fewer studies have examined the accuracy of self-reported height and weight when they are used to determine BMI categories [7-10]; yet, BMI categories are routinely used in studies of health outcomes [3]. Many of these studies showed significant differences in allocation to BMI categories based on self-reported versus measured height and weight, thus biasing relative risks of diseases associated with increasing BMI [3-8].

In women, bias in self-reported height and weight may occur due to social desirability, cultural or demographic characteristics or health characteristics (such as pregnancy or osteoporosis) [9]. In general, women tend to under-report weight more than men [2], while men tend to over-report height more than women [6]. It is important to examine the potential threats to accuracy particular to women since under- or over-reporting may affect the prevalence and categorization of BMI differently among women than among men. Understanding of sources of bias among women is important in planning and interpreting epidemiological studies based on self-reported height and weight. Mis-reporting could potentially reduce the utility of self-report upon which many women's health studies rely.

The primary aim of this study was to examine the strength of agreement between measured BMI categories and BMI categories based on self-reported heights and weights among women. Secondly, we examined potential biases among women who had discordant responses. For these purposes, we utilized self-reported and measured heights and weights in women from the 1999-2004 National Health and Nutrition Examination Survey (NHANES), a nationally representative sample of US adults.

Methods

National Health and Examination Survey, 1999-2004

Beginning in the 1960s, NHANES collected interview and physical examination data to assess the health and nutritional status of adults and children in the United States. In 1999, NHANES began to continuously survey a nationally representative sample of about 5,000 persons each year, and over the subsequent 5 years, 1999-2004, NHANES collected data on 31,123 subjects, including 8,970 women, age 18 or older. These participants were asked to self-report their height and weight along with a series of demographic, SES, health and health care questions (See Tables 1 and 2). On the same day as the survey, height and weight were subsequently measured using balance beam scales and a calibrated stadiometer as part of a standardized physician examination administered by the highly trained medical personnel in the Mobile Examination Center. The methods and study design for the program have been previously described [10].

The sample selection criteria includes women, age 18 or older, who provided complete interview responses, namely: self-reported height and weight, race, education, annual household income, marital status, number of live births, pregnancy, smoking history, health status on a five point scale, whether they have ever been diagnosed with osteoporosis, whether they have health insurance, and whether they visited their physician in the last 12 months. After exclusion of participants with missing anthropometric measurements, possibly due to amputations or weight above the limits of the scale, 200 kg, the sample size was reduced by 10.6% or 954 respondents. Missing interview responses further reduced the sample size by 17.2% or 1,382 respondents, resulting in an analytical file of 6,634 non-pregnant and pregnant women (See Table 1).

Given that pregnancy has an impact on body weight and may impact self-perception of body weight, we conducted a separate descriptive analysis of the pregnant subsample. Based on self-report and urine test results, 10.9% or 724 respondents were identified as pregnant at time of

the NHANES interview. Further analyses were conducted using the subsample of 5,910 non-pregnant respondents.

Over the 5-year period, 1999-2004, NHANES over-sampled low-income individuals, adolescents 12-19 years of age, individuals 60 years of age and older, African Americans, and Mexican Americans. The regression models include binary variables for these groups to address this over-sampling. Sampling weights were not applied because they were not constructed for the examination of non-pregnant women. We further recognize that the estimates are only generalizable to United States populations, from 1999 through 2004, with characteristics similar to those described in Table 2.

BMI Categories

Self-reported measurements of weight and height were collected in pounds, feet and inches, and were converted to kilograms and meters for the calculation of BMI, which is the ratio of weight in kilograms over height in meters squared (kg/m^2). We used pre-specified cutpoints of 18.5, 25, 30 and 40 kg/m^2 to divide respondents into BMI categories: underweight, normal, overweight, obese, and morbidly obese [11] (Table 2).

Statistical Analysis

Accuracy of a proxy measure may be assessed by its correspondence to a gold standard (e.g., probability of agreement) and the extent of bias in absence of agreement (e.g., likelihood of under-reporting relative to over-reporting). Optimally, BMI categories are based on height and weight measured through physical examination (e.g., balance beam scales and calibrated stadiometer by trained personnel), where agreement is defined as the equivalence between this gold standard and a proxy categorization using self-reported height and weight.

The interpretation of agreement between a gold standard and proxy measures depends on the likelihood of agreement at random. For example, if two thirds of an anorexic population is underweight and 75% randomly self-report underweight, then by randomness alone, half ($66\% \times 75\%$) should agree as underweight and a twelfth should agree ($33\% \times 25\%$) as not underweight; therefore the probability of agreement at random is 58.3% ($66\% \times 75\% + 33\% \times 25\%$) without any association between the proxy and gold standard measure. Formally, the probability of agreement at random over K categories is:

$$\text{Probability of Agreement at Random} = \sum_{k=1}^K \text{Pr}(\text{category}=k) * \text{Pr}(\text{proxy}=k) \quad (1)$$

Cohen's kappa is the probability of agreement adjusted for the probability of agreement at random, and this adjusted probability will be referred to as kappa for the remainder of the manuscript.

$$\text{Kappa} = \frac{\text{Probability of Agreement} - \text{Probability of Agreement at Random}}{1 - \text{Probability of Agreement at Random}} \quad (2)$$

If kappa equals one, the proxy and the gold standard are in perfect agreement. If the proxy and the gold standard are positively associated, kappa is positive. In rare cases, the proxy is negatively associated with the gold standard and kappa is less than zero. As reference, Landis and Koch put forth the labels for the strength of agreement based on kappa: Poor (<0.00), Slight ($0.00-0.20$), Fair ($0.21-0.40$), Moderate ($0.41-0.60$), Substantial ($0.61-0.80$) and almost Perfect ($0.81-1.00$) [12].

Kappa Regression

In addition to adjusting for random agreement, kappa has been expanded to assess the association between strength of agreement and respondent characteristics. Lipsitz et al. developed a practical linear regression approach for kappa [13]. Using their regression techniques, we examined potential threats to accuracy related to demographic, SES, health and behavior characteristics, in terms of differences in strength of agreement and bias. While some findings, such as those relating to SES, may be generalizable to men, our study exposes which subpopulations of women are most likely to exhibit inaccuracies in self-reporting.

The kappa regression procedure requires the estimation of individual-specific probabilities of agreement at random (see Eq. 1). Using multinomial logistic regression models, we predicted the probability of each BMI category (i.e., $\Pr(\text{category} = k|X)$) and the probability of each self-reported BMI category (i.e., $\Pr(\text{proxy} = k|X)$) (see Eq. 1), where the independent variables (X) are the characteristics that may influence the strength of agreement (Table 1). These two sets of predicted probabilities, one for the measured categories and another for the self-reported categories, are multiplied and aggregated for each respondent to compute the respondent-specific adjustment term, the probability of agreement at random. As shown in Eq. 2, the dependent variable of a kappa regression is the binary variable for agreement adjusted by the predicted probability of agreement at random.

The linear regression was estimated using ordinary least squares. Because of the dependence on auxiliary multinomial logistic regressions, estimated for the prediction of the adjustment term, the standard errors and P -values produced by ordinary least squares were underestimated. Instead of deriving a maximum likelihood approach that incorporates the linear and two logistic components, we bootstrap the coefficients, re-sampling the analytical sample 1,000 times and applying the percentile approach for the P -values [14,15]. Table 3 describes the changes in the kappa score associated with the respondent characteristics variables and the statistical significance of these changes based on the bootstrap results.

In addition to assessing agreement, we assess potential threats to accuracy, specifically, what are the odds of under-reporting among women relative to the risk of over-reporting? If this odds ratio is one, increased frequency of discordant respondents will not introduce bias. To estimate the ratio odds of under-reporting, we removed the respondents whose self-reported BMI categories agreed with the measured categories ($N = 4,670$). By construction, concordant responses do not introduce bias. Furthermore, we removed discordant respondents who were underweight ($N = 54$) or morbidly obese ($N = 129$), because their discordant responses are unidirectional by construction. For the remaining discordant respondents ($N = 1,057$), they may over- or under-report their category. The ratio odds of under-reporting versus over-reporting, shown in Table 4, were estimated by binomial logistic regression. We further describe the probability of under-reporting (or one minus the probability of over-reporting) in the text to facilitate discussion.

In both the logistic and kappa regression models, the null case represents a white, non-Hispanic woman, age 26 to 35, who is a single, insured, high school graduate with more than \$20,000 in annual household income. Furthermore, this person never smoked, never experienced a live birth, is in very good health, is not pregnant, and has visited her physician at least once in the last 12 months. This case was selected based on median or modal values of the variables (Table 1).

Database management was conducted using SAS 9.1 and the resulting analytical samples were examined using STATA MP 9.2. The study was reviewed and approved by the University of Wisconsin Institutional Review Board, which considered the study exempt due to its use of publicly available data sets (45 CFR 46.101(b) (4)).

Results

Descriptive Data

In this study, 5,910 non-pregnant women were included in the sample and Table 1 illustrates the characteristics of the sample. The majority of respondents (90%) report at least one physician visit in the last 12 months, which suggests some previous clinical measurement of height and weight within the year. Table 2 illustrates the frequency of self-reported BMI vs. measured BMI categories. Self-reported categories are equivalent to measured categories for 79% of non-pregnant women, with 21% of non-pregnant women being placed in the wrong BMI category when BMI category is based on self reported height and weight. For comparison, only 60% of the pregnant women had categorical agreement. After adjusting for random agreement, the kappa estimates are 0.705 (SE 0.008) for non-pregnant women and 0.443 (SE 0.022) for pregnant women. Based on conventional standards, we find substantial agreement among non-pregnant women, and moderate agreement among pregnant women for the overall sample.

Strength of Agreement

The strength of agreement between self-reported and measured BMI categories varied significantly according to age and race (Table 3). Agreement was greatest among white, non-Hispanic women, age 26-35, and decreased significantly among racial minorities and women older than 66 years of age. In terms of clinical significance, agreement was moderate or better across demographic subpopulations. SES, as categorized based on education and income, was unrelated to strength of agreement. Pregnancy significantly decreased the strength of agreement.

Among the behavioral and health characteristics, only the self-reported diagnosis of osteoporosis had a statistically significant effect on accuracy (Table 3). The finding that maternity, smoking history, and poor health were unrelated to the accuracy of self-reported categories is also noteworthy.

Access to health care through marriage, insurance and annual visits was largely unrelated to agreement. Respondents who had a physician visit in the last 12 months (i.e., access to clinical measurement) had stronger agreement than respondents who had not. However, the absence of an annual visit decreased kappa by only 0.085.

Likelihood of Under-Reporting

To characterize bias, we examined the sample of discordant responses from non-pregnant women who were either normal weight, overweight, or obese ($N = 1,057$). Overall, 76.5% of this sub-sample under-report their weight, which is lower than the 92.6% under-reporting found among pregnant women. Optimally, these probabilities would be 50%. Based on this evidence, pregnant women are 21% (0.927/0.765) more likely to under-report their weight, which suggests that the self-reported BMI category is less accurate, in terms of agreement and bias, for pregnant women compared to non-pregnant women.

Table 4 shows the odds of under-reporting among discordant responders based on logistic regression. The probability of under-reporting among discordant non-pregnant women (i.e., odds/1 + odds) varied by age and race (Table 4). After adjusting for demographic, health, and SES characteristics, the probability of under-reporting by race and ethnicity indicator was 76% for white, non-Hispanic women, 63% for black women, 67% for Hispanic women, and 58% for other minorities. These probabilities are significantly different from 50% for white and Hispanic women; therefore white and Hispanic women are more likely to under-report than over-report. Under-reporting was also found across all age groups, particularly women, age

46-55. Demographically speaking, bias appears greatest among older white, non-Hispanic women.

Bias was also found to vary across SES categories based on education and income. Compared to discordant women with median SES and lower SES (No High School Diploma & Income under \$20,000), discordant women with some higher education are more likely to under-report their obesity (OR 2.03-5.12, $P < 0.0001$) (Table 4). Access to health care through marriage, insurance and annual visit was statistically unrelated to bias among discordant women.

Discussion

Based on our study of women in NHANES, BMI categories based on self-reported height and weight had substantial agreement with measured categories among non-pregnant women. Agreement between BMI category based on self-reported and measured height and weight was significantly related to age and race, with less agreement found for older, black and Hispanic women. Strength of agreement was unrelated to SES, access to care or health status. However, pregnancy significantly decreased strength of agreement.

In the subset of women with discordant responses, the majority under-reported their obesity category. In particular, we found an under-reporting bias among white, non-Hispanic women with some college education. Thus, concerns, in terms of accuracy, would be greatest for studies that rely on self-reported categories for women with these characteristics. However, even in this worst case scenario, BMI categories based on self-reported height and weight still demonstrate moderate agreement with measured categories.

Behavioral and health characteristics were largely unrelated to the accuracy of BMI categories based on self-reported height and weight, except to mention that women who annually visit a physician or are diagnosed with osteoporosis were more accurate in their reporting. A surprising finding was that the diagnosis of osteoporosis, a condition associated with a loss in height, improved agreement among older women. Women, age 56-65 years and diagnosed with osteoporosis, may more accurately report their BMI category due to access to a physician, due to better self monitoring of physiological changes, or due to greater awareness of a potentially latent conditions.

Our findings suggest that self-reported and measured BMI categories among pregnant women are in moderate agreement, and discordant responders largely under-report their category. The possibility that pregnant women can accurately assess their weight seems highly unlikely based on this evidence, and we suggest caution in the interpretation of such data. Further studies are warranted to address issues concerning recall bias and gestational development.

Our results were similar to many studies which found that women are more likely to under-report weight [6,8,16-18]. Similar to Nyholm, we found that age is related to bias in self-reported BMI [9]. However, because we examined BMI categories, not BMI in its continuous form, we were able to examine the extent to which this bias might threaten categorical agreement. We found that while differences between self-reported and clinically measured values are statistically significant, varying by age and race, self-reported BMI categories show substantial agreement with clinical measures across all age and race subgroups.

In addition to evidence supporting the use of self-reported BMI categories, the results contribute to our understanding of clinical effects. Our results suggest that a diagnosis of osteoporosis increases the accuracy of self-reported BMI categories, particularly in women age 55-76 and presumably because these women are more aware of their true height. It is interesting that women with previous diagnoses of osteoporosis are more likely to report accurately as many geriatric studies have found self-reports to be less reliable in older women.

Wada and colleagues found no significant differences in measured versus self-reported BMI in males or females with hypertension or hyperlipidemia, but did find a significant difference in diabetics, who more accurately self-reported BMI [19].

Prior to our study, no study has examined the agreement between self-reported and measured BMI categories beyond the probability of agreement at random (i.e., Cohen's kappa). Brunner Huber found that self-reported height and weight measures classified 84% of women of reproductive age into appropriate BMI categories [7]. While this article reports percentage agreement, it excluded a measure of agreement adjusted for agreement at random (i.e., Cohen's kappa) from the analysis. Still, Cohen's kappa (0.77) can be computed using the data found in Table 6 of the article. Some studies have simplified the five BMI categories into a binary variable (e.g., obese and non-obese) and examined sensitivity and specificity. This simplification may facilitate explanation, because it describes the probability that an obese patient will be categorized as obese based on self-reported weight and height (true-positive rate). However, it loses descriptive power by equating underweight, normal and overweight individuals as well as equating obese and morbidly obese individuals. Furthermore, using sensitivity and specificity does not account for agreement at random. In the previous example of an anorexic population, kappa would be zero, but sensitivity of underweight (i.e., probability of agreement conditional on underweight) would be 75%, which may be erroneously interpreted as a strong association.

A key limitation of the study is that NHANES participants may have known they were going to be weighed when they consented for the study, which may have decreased tendencies to under-report weight. Any comparison of self-reported and clinically reported height and weight requires consent; therefore, this would be susceptible to experimental influences in addition to social desirability bias.

Conclusions

Our results suggest substantial agreement between self-reported and measured categories, except for women who are pregnant, who have not seen a physician in the last year or are above age 75. The use of self-reported height and weight, recoded as BMI categories to evaluate obesity, can be affected by misclassification and may lead to underestimation of the prevalence of the overweight and obese in certain populations. Such bias may be greater in studies that examine well-educated or largely white populations compared to studies that examine minority obesity. This is important in light of increasing obesity trends in the US and the ongoing need for obesity surveillance and treatment.

Acknowledgements

The authors thank the staff members, Mary Palmer, Kelly Muenzenberger, Angeline Vanto, and Kevin Benish, for their support of this project.

References

1. DHHS-CDC. Morbidity and Mortality Weekly Report. 2006. State-specific prevalence of obesity among adults—United States, 2005; p. 985-1012.
2. Kuczmarski MF, Kuczmarski RJ, Najjar M. Effects of age on validity of self-reported height, weight, and body mass index: Findings from the Third National Health and Nutrition Examination Survey, 1988-1994. *Journal of the American Dietetic Association* 2001;101:28-34. [PubMed: 11209581]doi: 10.1016/S0002-8223(01)00008-6
3. Must A, Spandano J, Coakley E, Field A, Colditz G, Dietz W. The disease burden associated with overweight and obesity. *Journal of the American Dietetic Association* 1999;282:1523-1529.

4. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: Sensitivity and specificity using self-reported weight and height. *Epidemiology (Cambridge, Mass)* 1990;1:146–152.doi: 10.1097/00001648-199003000-00011
5. Schmidt MI, Duncan BB, Tavares M, Polanczyk CA, Pellanda L, Zimmer PM. Validity of self-reported weight—a study of urban Brazilian adults. *Revista de Saude Publica* 1993;27:271–276. [PubMed: 8209159]doi:10.1590/S0034-89101993000400007
6. Bostrom G, Diderichsen F. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *International Journal of Epidemiology* 1997;26:860–866. [PubMed: 9279620]doi:10.1093/ije/26.4.860
7. Brunner Huber LR. Validity of self-reported height and weight in women of reproductive age. *Maternal and Child Health Journal* 2007;11:137–144. [PubMed: 17066316]doi:10.1007/s10995-006-0157-0
8. Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutrition* 2002;5:561–565. [PubMed: 12186665]doi: 10.1079/PHN2001322
9. Nyholm M, Gullberg B, Merlo J, Lundqvist-Persson C, Rastam L, Lindblad U. The validity of obesity based on self-reported weight and height: Implications for population studies. *Obesity (Silver Spring, Md.)* 2007;15:197–208.
10. Centers for Disease Control and Prevention NCHS. National Health and Nutrition Examination Survey Data. US DHHS, Centers of Disease Control and Prevention; Hyattsville, MD: 19992004.
11. McDowell MA, Hughes JP, Borrud LG. Health characteristics of U.S. adults by body mass index category: Results from NHANES 1999–2002. *Public Health Report* 2006;121:67–73.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174. [PubMed: 843571]
13. Lipsitz AR, Williamson J, Klar N, et al. A simple method for estimating a regression model for κ between a pair of raters. *Journal of Royal Statistical Society Series A* 2001;64(3):449–465.doi: 10.1111/1467-985X.00213
14. Shoukri MM, Mian IU. Maximum likelihood estimation of the kappa coefficient from bivariate logistic regression. *Statistics in Medicine* 1996;15:1409–1419. [PubMed: 8841651]doi:10.1002/(SICI)1097-0258(19960715)15:13<1409::AID-SIM269>3.0.CO;2-N
15. Efron, BT.; Tibshirani, RJ. An introduction to the bootstrap (Monographs on statistics and applied probability). Vol. 1st ed.. Chapman & Hall/CRC; 1993.
16. Niedhammer I, Bugel I, Bonenfant S, Goldberg M, Leclerc A. Validity of self-reported weight and height in the French GAZEL cohort. *International Journal of Obesity and Related Metabolic Disorders* 2000;24:1111–1118. [PubMed: 11033979]doi:10.1038/sj.ijo.0801375
17. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *Journal of Epidemiology and Community Health* 2000;54:143–148. [PubMed: 10715748]doi:10.1136/jech.54.2.143
18. Madrigal H, Sanchez-Villegas A, Martinez-Gonzalez MA, et al. Underestimation of body mass index through perceived body image as compared to self-reported body mass index in the European Union. *Public Health* 2000;114:468–473. [PubMed: 11114759]doi:10.1016/S0033-3506(00)00390-5
19. Wada K, Tamakoshi K, Tsunekawa T, et al. Validity of self-reported height and weight in a Japanese workplace population. *International Journal of Obesity* 2005;29:1093–1099. [PubMed: 15925952] doi:10.1038/sj.ijo.0803012

Table 1
 Characteristics of non-pregnant respondents ($N = 5,910$)

	%	<i>N</i>
Age at interview in years		
18-25	0.18	1082
26-35	0.14	817
36-45	0.17	991
46-55	0.14	835
56-65	0.14	825
66-75	0.12	703
Over 75	0.11	657
Race and ethnicity		
White, non-hispanic	0.50	2943
Black	0.20	1207
Hispanic	0.26	1562
Other	0.03	198
Education		
Less than high school diploma	0.31	1828
High school diploma	0.26	1509
More than high school diploma	0.44	2573
Annual household income		
Under \$20,000	0.29	1697
Greater than or equal to \$20,000	0.71	4213
Marital status		
Married	0.44	2607
Not married	0.56	3303
Self-reported health status		
Fair or poor	0.22	1328
Excellent, very good, or good	0.78	4582
Birth history		
No live births	0.29	1722
One or more live births	0.71	4188
Smoking history		
Current smoker	0.17	1012
Former smoker	0.19	1122
Never smoked	0.64	3776
Access to care		
No physician visits in the last 12 months	0.10	581
Ever told you had osteoporosis	0.09	543
Uninsured at time of interview	0.18	1065

Table 2
 Non-pregnant respondents by self-reported and measured BMI categories^a

Measured BMI category	BMI category based on self-reported height and weight					Total
	Underweight	Normal	Overweight	Obese	Severely obese	
Underweight	85	54	0	0	0	139
Normal	45	1,811	104	5	2	1,967
Overweight	0	390	1,275	103	3	1,771
Obese	0	17	357	1,248	31	1,653
Severely obese	0	1	0	128	251	380
Total	137	2,375	1,793	1,536	311	5,910

^aDiagonal cells (N = 4670; 79%) represent agreement between self-reported and measured BMI categories

Table 3
Strength of agreement between self-reported and measured BMI categories based on Cohen's kappa

Characteristics of non-pregnant respondents	Change in kappa ^a	Standard error	P-value
Age in years			
18-25	-0.081	0.029	0.006
26-35 ^b	-	-	-
36-45	-0.035	0.027	0.242
46-55	-0.019	0.027	0.512
56-65	-0.064	0.029	0.030
66-75	-0.140	0.033	0.000
Over 75	-0.184	0.038	0.000
Race and ethnicity			
White, non-Hispanic ^b	-	-	-
Black	-0.061	0.022	0.004
Hispanic	-0.062	0.022	0.002
Other	-0.015	0.043	0.788
Education and annual household income categories			
No high school diploma & income under \$20,000	-0.047	0.030	0.124
No high school diploma & income greater than or equal to \$20,000	-0.005	0.027	0.862
High school diploma & income under \$20,000	0.010	0.034	0.780
High school diploma & income greater than or equal to \$20,000 ^b	-	-	-
More than high school & income under \$20,000	0.009	0.035	0.830
More than high school & income greater than or equal to \$20,000	0.020	0.021	0.356
Access to care			
Married	0.025	0.018	0.164
Uninsured at time of interview	-0.027	0.024	0.270
No physician visits in the last 12 months	-0.085	0.029	0.002
Behavioral and health characteristics			
One or more live births	0.005	0.020	0.758
Current smoker	-0.033	0.022	0.114
Former smoker	0.021	0.021	0.322
Fair or poor health status	0.023	0.020	0.256
Age 56 to 65 and ever diagnosed with osteoporosis	0.125	0.050	0.024
Age 66 to 75 and ever diagnosed with osteoporosis	0.112	0.053	0.024
Age 76 or older and ever diagnosed with osteoporosis	0.007	0.054	0.880

^aThe predicted kappa value for the null case is 0.775 with a standard error of 0.032

^bThe null kappa estimate (0.775) predicts substantial strength of agreement

Table 4
Response bias based on odds of underreporting versus overreporting

Characteristics of non-pregnant respondents with discordant reporting ^a	Odds ratio	Standard error	P-value
Age in years			
18-25	0.873	0.270	0.662
26-35 ^b	-	-	-
36-45	0.765	0.223	0.358
46-55	0.543	0.166	0.046
56-65	0.764	0.245	0.401
66-75	1.285	0.439	0.463
Over 75	1.443	0.548	0.334
Race and ethnicity			
White, non-Hispanic ^b	-	-	-
Black	0.529	0.113	0.003
Hispanic	0.636	0.133	0.031
Other	0.431	0.182	0.046
Education and annual household income categories			
No high school diploma & income under \$20,000	1.708	0.456	0.045
No high school diploma & income greater than or equal to \$20,000	0.841	0.198	0.463
High school diploma & income under \$20,000	1.021	0.325	0.949
High school diploma & income greater than or equal to \$20,000 ^b	-	-	-
More than high school & income under \$20,000	5.123	2.357	0.000
More than high school & income greater than or equal to \$20,000	2.029	0.460	0.002
Access to care			
Married	1.049	0.183	0.785
Uninsured at time of interview	1.227	0.243	0.301
No physician visits in the last 12 months	0.963	0.223	0.870
Behavioral and health characteristics			
One or more live births	1.223	0.229	0.282
Current smoker	0.810	0.168	0.310
Former smoker	0.735	0.162	0.161
Fair or poor health status	0.919	0.178	0.663
Age 56-65 and ever diagnosed with osteoporosis	1.734	1.408	0.498
Age 66-75 and ever diagnosed with osteoporosis	1.050	0.727	0.944
Age 76 or older and ever diagnosed with osteoporosis	1.967	1.183	0.261

^a Respondents whose self-reported categories agree with clinical measures were removed for this analysis to allow comparison of underreporting and overreporting respondents. Furthermore, under-weight and severely obese respondents were removed, because disagreements are unidirectional by construction. The resulting sub-sample $N = 1,057$ respondents

^b The predicted odds of underreporting for the null case is 2.63