

Practice of Epidemiology

Bias in Hazard Ratios Arising From Misclassification According to Self-Reported Weight and Height in Observational Studies of Body Mass Index and Mortality

Katherine M. Flegal*, Brian K. Kit, and Barry I. Graubard

* Correspondence to Dr. Katherine M. Flegal, Prevention Research Center, Stanford University School of Medicine, Medical School Office Building, 1265 Welch Road, Mail Code 5411, Stanford, CA 94305-5411 (e-mail: kmflegal@stanford.edu).

Initially submitted July 13, 2016; accepted for publication February 6, 2017.

Misclassification of body mass index (BMI) categories arising from self-reported weight and height can bias hazard ratios in studies of BMI and mortality. We examined the effects on hazard ratios of such misclassification using national US survey data for 1976 through 2010 that had both measured and self-reported weight and height along with mortality follow-up for 48,763 adults and a subset of 17,405 healthy never-smokers. BMI was categorized as <22.5 (low), 22.5–24.9 (referent), 25.0–29.9 (overweight), 30.0–34.9 (class I obesity), and ≥ 35.0 (class II–III obesity). Misreporting at higher BMI categories tended to bias hazard ratios upwards for those categories, but that effect was augmented, counterbalanced, or even reversed by misreporting in other BMI categories, in particular those that affected the reference category. For example, among healthy male never-smokers, misclassifications affecting the overweight and the reference categories changed the hazard ratio for overweight from 0.85 with measured data to 1.24 with self-reported data. Both the magnitude and direction of bias varied according to the underlying hazard ratios in measured data, showing that findings on bias from one study should not be extrapolated to a study with different underlying hazard ratios. Because of misclassification effects, self-reported weight and height cannot reliably indicate the lowest-risk BMI category.

body mass index; body weight; epidemiologic methods; mortality; NHANES; obesity; overweight; self-report

Abbreviations: BMI, body mass index; NHANES, National Health and Nutrition Examination Survey.

Studies of weight and mortality commonly assess weight using body mass index (BMI, calculated as weight (kg)/height (m)²), a form of weight adjusted for height. A number of large studies of BMI and mortality have calculated BMI from self-reported weight and height (1–4). However, BMI calculated from self-reported weight and height (“self-reported BMI”) is systematically biased relative to BMI calculated from objectively measured weight and height (“measured BMI”) (5). Errors are systematic rather than random in part because of the tendencies for greater underreporting of weight at high weights and greater overreporting of height at shorter heights (6). Additional factors related to error in reporting include age, gender, and race (7–9).

The rate of misclassification into the wrong BMI categories, relative to measured BMI, is often quite high. For example, Spencer et al. (10) found that almost 30% of men classified as normal weight by self-report were classified as overweight or obese by measured data. One effect of misclassification is to bias the estimated prevalence of obesity downward when self-

reported data are used (11–16). However, misclassification affecting an exposure variable, such as BMI category, can also result in biased estimates of the relationship of the exposure to an outcome such as mortality (17). Therefore misclassification arising from self-reported weight and height may bias hazard ratios in studies of BMI and mortality.

Comparisons of self-reported with measured BMI typically find small mean differences and very high correlations, often in the range of 0.95–0.99. However, despite what has sometimes been stated (18), this does not demonstrate that self-reported BMI introduces little bias. Correlation does not assess agreement between 2 methods (19). Neither correlations nor mean differences address the bias in hazard ratios arising from misclassification.

Hazard ratios for the overweight and obesity categories of self-reported BMI tend to be higher than those for the overweight and obesity categories of measured BMI (20, 21). The systematic errors characteristic of self-reported data would be expected to bias hazard ratios for obesity upwards relative to

measured BMI because of underreporting at the highest BMI levels (22–24). The occurrence of this phenomenon has been documented in a number of studies (25–27). Similar predictions have been noted in the context of dietary data, alcohol-intake data, and occupational-exposure data, where underreporting of exposure at high levels would be expected to increase hazard ratios (28–30).

Several previous studies (21, 31) using data from the National Health and Nutrition Examination Survey (NHANES) have examined the overall effects on hazard ratios of errors in BMI due to self-reported weight and height. Here we extend these observations to provide examples of specific types of BMI misclassification that can affect hazard ratios, including misclassifications that affect the reference BMI category. We restrict our attention to the methodological issues regarding comparisons of self-reported and measured BMI values and their association with mortality. A 2009 Science Advisory from the American Heart Association noted that there are few data to directly evaluate the magnitude and direction of bias due to self-reported weight and height and commented that such studies “would add greatly to the literature” (32, p. 3266).

METHODS

Data in this study come from the series of NHANES surveys conducted by the National Center for Health Statistics. In each survey, a different nationally representative cross-sectional sample of the US population was interviewed and examined. Numerous other articles (33–43) have addressed broader issues of adiposity and mortality in the NHANES data.

We used baseline data from NHANES II (1976–1980), NHANES III (1988–1994), and the Continuous NHANES (1999–2010) (44–46), with linked mortality data through 2011. All surveys included a household interview followed within a few weeks by a standardized physical examination in a mobile examination center. Questions about weight and height were included in the household interview. Weight and height were subsequently measured in the examination center using standardized methods and equipment. At the time of interview, respondents agreed to participate in the physical examination and were informed that it would include height and weight measurements.

There were 48,781 adult participants who had complete data for measured and self-reported weight and height, smoking status, racial/ethnic group, alcohol consumption, and educational level. After excluding 18 participants who had a difference of ≥ 20 BMI units between self-reported and measured BMI, the final analytical data set included 48,763 participants with 13,058 deaths.

For the purpose of these analyses, BMI was categorized as <22.5 (low), 22.5–24.9 (referent), 25.0–29.9 (overweight), 30.0–34.9 (class I obesity), and ≥ 35.0 (class II–III obesity). A BMI of 22.5–24.9 was used as the reference category because this is frequently selected as the reference category in large studies (17). For the current analyses, a BMI of <22.5 will be referred to as “low BMI.”

We used sex-specific hazard ratios for all-cause mortality, computed from Cox proportional hazards models, with age as the timeline and adjusted for survey, smoking status, racial/ethnic

group, and alcohol consumption (36). We calculated hazard ratios for BMI categories using measured BMI and again using self-reported BMI. We repeated the analyses for a restricted data set with 17,405 participants that included only self-reported never-smokers who had been examined below the age of 70 years and who reported no history of heart disease or cancer. These restrictions were chosen because they are similar to those used in a number of large studies of weight and mortality (2–4).

Misclassifications between categories can affect the comparison between hazard ratios from measured versus self-reported data in a variety of ways. For example, the hazard ratio for overweight relative to the reference category may be affected by misclassifications that affect the reference category, misclassifications that affect the overweight category, or both. Because we used 5 BMI categories, there were 20 possible pairs of misclassifications between categories. However, several of those pairs never occurred in the data or occurred so infrequently as to be negligible. We considered only misclassifications of 1 category higher or lower than the correct measured BMI category, leading to 8 possible misclassifications (low to referent, referent to low, referent to overweight, overweight to referent, overweight to obesity class I, obesity class I to overweight, obesity class I to obesity class II–III, obesity class II–III to obesity class I). Over 99% of the sample fell within this range from 1 category lower to 1 category higher.

To isolate the effect of misclassification from a specific BMI category to another specific category, we substituted self-reported BMI for measured BMI just for the misclassified participants and recalculated hazard ratios, retaining measured BMI for the other categories. For example, for misclassification from the low category to the reference category, we substituted self-reported BMI for measured BMI for those who were in the low category by measured BMI but in the reference category by self-reported BMI and then recalculated the hazard ratios. We repeated this for every possible combination of the 8 misclassifications, substituting self-reported BMI for measured BMI for that particular combination, resulting in 256 combinations. For presentation purposes, we selected the 5 combinations of individual misclassifications that could affect each of the 5 BMI categories, as shown in Web Table 1 (available at <https://academic.oup.com/aje>). The highest and lowest BMI categories were affected by only 2 misclassifications; the other BMI categories were each affected by 4 misclassifications. Each individual misclassification affects 2 BMI categories.

These methods were designed to elucidate how and why the point estimates of hazard ratios numerically change under the observed misclassification of self-reported BMI. This objective differs from that of Keith et al. (47), whose objective was primarily to determine whether the statistical significance of the results would differ when using measured BMI compared with self-reported BMI, and from that of Preston et al. (48), who considered the bias due to categorization and self-report relative to measured data and their effects on attributable fractions.

Analyses were conducted with PC SAS, version 9.4 (SAS Institute, Inc., Cary, North Carolina), and SUDAAN, version 11.0.1 (RTI International, Research Triangle Park, North Carolina). All analyses used sample weights and took into account the clustering and the sample design.

RESULTS

In both the full sample and the restricted sample of healthy never-smokers, self-reported and measured BMI were highly correlated, with correlation coefficients of 0.96 or 0.97 between self-reported BMI and measured BMI. The mean differences were small, with a mean difference of 0.4 units for men and 0.7 units for women. Details of descriptive data are shown in Table 1. These observations do not assess classification by BMI.

Misclassification rates are shown in Figure 1, which shows (for each category of self-reported BMI) the proportions that were in the same or different measured BMI categories. For instance, of men classified into the reference BMI category by self-reported weight and height, 64% were correctly classified, 22% fell into the next higher BMI category by measured weight and height, and 13% fell into the next lower BMI category. Misclassification rates for the restricted sample were similar to those for the full sample. Details of the cross-classification by self-reported and measured BMI groupings are shown in Web Table 2. The percentage of overall agreement and Cohen's kappa measure of interrater agreement are shown in Web Table 3. Percent agreement overall ranged from 76.5% to 78.1%, and values of kappa ranged from 0.69 to 0.71.

Effects of data collection methods and sample restrictions

The overall sex-specific hazard ratios for self-reported and measured BMI for the full sample and the restricted sample are shown in Table 2. For both men and women in the full sample, the hazard ratios tended to be slightly higher at the highest BMI category when self-reported BMI was used. For men in the restricted sample, the hazard ratios calculated using self-reported BMI were higher for all BMI categories than the hazard ratios calculated using measured BMI. When measured BMI was used, the hazard ratios for low BMI and

overweight BMI were both lower than 1; when self-reported BMI was used, the direction of the hazard ratios was reversed and both were greater than 1. For women in the restricted sample, unlike the results for men, the hazard ratios calculated using self-reported BMI were lower for all BMI categories than the hazard ratios calculated using measured BMI.

Effects of misclassifications between combinations of specific BMI categories

The above results are for all misclassifications taken together. We carried out a detailed examination of how the hazard ratios were affected by misclassifications between specific BMI categories. We calculated results separately for 256 combinations of misclassification effects. Table 3 for the full sample and Table 4 for the restricted sample summarize the effects on hazard ratios of the combinations that affect each BMI category. For example, for men in the full sample, the second line of Table 3 shows the combined impact on hazard ratios of the 2 misclassifications that affect the low self-reported BMI category (misclassification from low to referent and misclassification from referent to low). These misclassifications increased the hazard ratio for the low BMI category from 1.28 to 1.30, with similarly small increases for the hazard ratios in the other BMI categories. The third line shows the combined impact on hazard ratios of the group of the 4 misclassifications that affect the reference category (low to referent, referent to low, overweight to referent, referent to overweight). These misclassifications increased the hazard ratio for the low BMI category from 1.28 to 1.30, had no effect on the hazard ratio for the overweight category, and slightly increased the hazard ratio for the obesity BMI categories.

In the full sample, misclassifications that affected the low, reference, or overweight categories had little impact, as shown in Table 3. For men, misclassifications that affected class I or class II–III obesity increased the hazard ratios for those categories. However, for women in the full sample, those

Table 1. Descriptive Data From National Health and Nutrition Examination Survey II (1976–1980), III (1988–1994), and Continuous (1999–2010), United States

Sample and BMI Ascertainment Method	Men				Women			
	No. of Participants	Mean	Median (IQR)	Correlation of Measured and Self-Reported BMI	No. of Participants	Mean	Median (IQR)	Correlation of Measured and Self-Reported BMI
Full sample	24,136			0.96	24,627			0.97
Measured ^a		27.9	27.2 (6.42)			28.0	26.5 (8.87)	
Self-reported ^b		27.6	26.7 (6.00)			27.2	25.8 (8.22)	
Restricted sample ^c	6,999			0.97	10,406			0.97
Measured ^a		28.3	27.4 (6.51)			28.0	26.4 (9.14)	
Self-reported ^b		28.0	27.2 (6.16)			27.3	25.8 (8.36)	

Abbreviations: BMI, body mass index; IQR, interquartile range.

^a Measured BMI was calculated from measured height and measured weight as weight (kg)/height (m)².

^b Self-reported BMI was calculated from self-reported weight and self-reported height as weight (kg)/height (m)².

^c The restricted sample was limited to participants who reported never smoking, who reported no history of heart disease or cancer, and who were examined when younger than age 70 years.

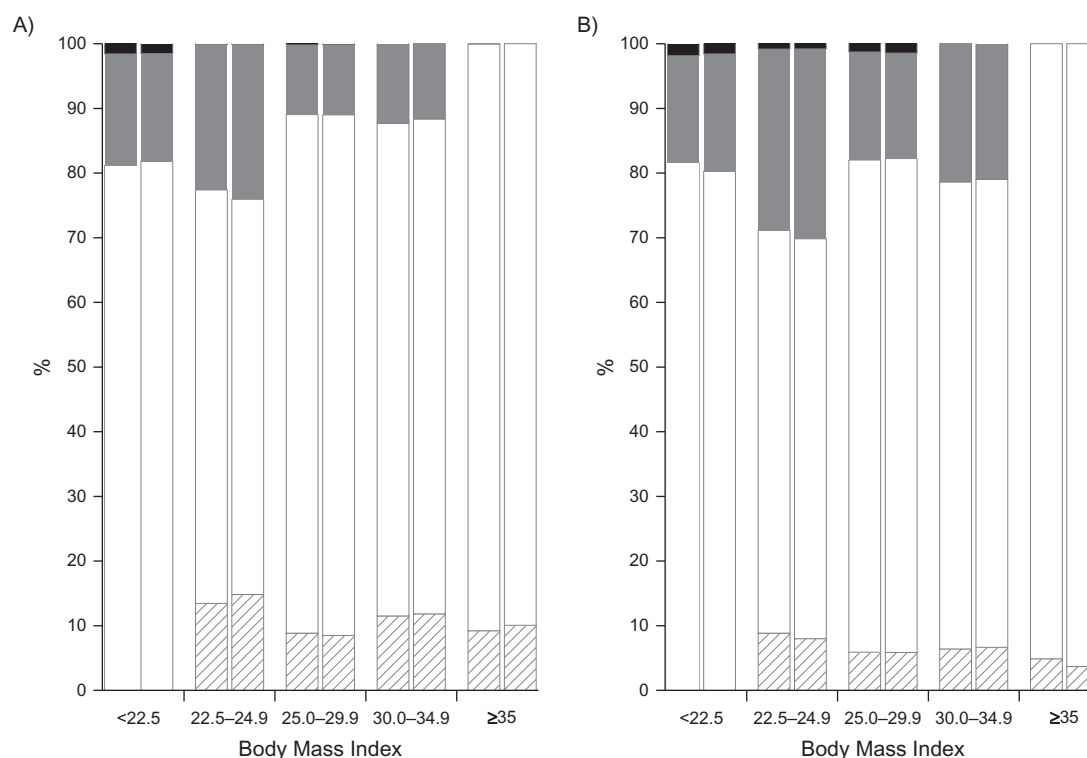


Figure 1. Percentage of sample within categories of self-reported body mass index (BMI) that were classified into the same, higher, or lower categories of measured BMI, National Health and Nutrition Examination Survey (NHANES) II (1976–1980), NHANES III (1988–1994), and Continuous NHANES (1999–2010), United States. A) Men. B) Women. Measured BMI was calculated from measured height and weight. Self-reported BMI was calculated from self-reported weight and height. The x-axis BMI categories are according to self-report. Within BMI categories, the first bar is for the full sample, and the second bar for the restricted sample. White: measured BMI category was the same as self-reported BMI category; striped: measured BMI category was lower than self-reported BMI category; gray: measured BMI category was 1 category higher than self-reported BMI category; black: measured BMI category was 2 categories higher than self-reported BMI category.

misclassifications had almost no effect. The changes were small in all BMI categories.

The results were different for the restricted sample of healthy never-smokers, as shown in Table 4. For men in the restricted sample, misclassifications that affected class I or class II–III obesity increased the hazard ratios for the highest BMI category but had almost no effect on the hazard ratios for any other BMI category. However, misclassifications that affected the low, reference, or overweight categories increased the hazard ratios for all BMI categories, and misclassifications affecting the reference or overweight categories changed the direction of the hazard ratio for the overweight category.

For women in the restricted sample, in contrast, the use of self-reported BMI decreased the hazard ratios in all BMI categories. These results appeared initially to be opposite to those for the men. However, as for men, misclassifications that affected class I or class II–III obesity increased the hazard ratios for the highest BMI categories with little or no effect on the lower BMI categories. Misclassifications affecting the low, reference, or overweight categories produced the opposite effects, reducing the hazard ratios in all BMI categories. Further investigation suggested that these decreases in hazard ratios were due entirely to the single misclassification from the low category into the reference category. Because the hazard

ratio for the low category (1.37) was higher than that for the reference category, misclassification from the low category to the reference category increased the apparent risk in the reference category, which then reduced the hazard ratio for all other BMI categories.

Effects of individual BMI misclassifications

Tables 3 and 4 show the effects of groups of misclassifications but do not show the effects of each individual misclassification. An example of how cumulative individual misclassifications changed the direction of the hazard ratio for the overweight category for men in the restricted sample is displayed in Table 5. The use of self-reported data changed the direction of the hazard ratio for overweight from 0.85 with measured BMI to 1.24 for self-reported BMI and increased the hazard ratio for the class II–III obesity category from 1.86 to 2.88. Only 4 individual misclassifications were enough to change the direction. With no misclassifications, the hazard ratio was 0.85. Two-way misclassification between the lower risk overweight category and the higher risk reference category increased the hazard ratio from 0.85 to 1.08. Two further misclassifications increased the risk in the overweight category and decreased the risk in the reference category, thereby increasing the hazard ratio to 1.24.

Table 2. Mortality Hazard Ratios^a for Measured or Self-Reported Categories of Body Mass Index According to Sex in the Full and Restricted^b Samples From National Health and Nutrition Examination Survey II (1976–1980), III (1988–1994), and Continuous (1999–2010), United States

Sample and BMI Ascertainment Method	BMI Category							
	Low (<22.5)		Overweight (25–29.9)		Class I Obesity (30–34.9)		Class II–III Obesity (≥35)	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
<i>Men</i>								
Full sample								
Measured ^c	1.28	1.13, 1.45	0.94	0.86, 1.02	1.08	0.95, 1.22	1.70	1.42, 2.03
Self-reported ^d	1.30	1.16, 1.44	0.97	0.89, 1.05	1.07	0.94, 1.22	1.83	1.52, 2.20
Restricted sample								
Measured ^c	0.87	0.56, 1.36	0.85	0.62, 1.17	1.13	0.78, 1.64	1.86	1.12, 3.07
Self-reported ^d	1.26	0.82, 1.94	1.24	0.90, 1.70	1.26	0.84, 1.90	2.88	1.76, 4.72
<i>Women</i>								
Full sample								
Measured ^c	1.19	1.05, 1.34	1.09	0.97, 1.23	1.30	1.15, 1.48	1.65	1.40, 1.94
Self-reported ^d	1.18	1.07, 1.31	1.13	1.02, 1.24	1.37	1.21, 1.54	1.69	1.42, 2.02
Restricted								
Measured ^c	1.37	1.04, 1.81	1.21	0.96, 1.52	1.59	1.24, 2.03	1.94	1.49, 2.54
Self-reported ^d	1.15	0.91, 1.45	1.11	0.88, 1.39	1.48	1.14, 1.91	1.81	1.34, 2.46

Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio.

^a Hazard ratios from Cox proportional hazards models, with age as the timeline and adjusted for survey, smoking status, racial/ethnic group, and alcohol consumption. The reference category was BMI of 22.5–24.9.

^b The restricted sample was limited to participants who reported never smoking, who reported no history of heart disease or cancer, and who were examined when younger than age 70 years.

^c Measured BMI was calculated from measured height and measured weight as weight (kg)/height (m)².

^d Self-reported BMI was calculated from self-reported weight and self-reported height as weight (kg)/height (m)².

Detailed tables for each BMI group for the restricted sample, showing the sequential effects of each misclassification, are presented in Web Tables 4 (men) and 5 (women). In several cases, a version of what is known as the “Will Rogers phenomenon” (49–52) was apparent, whereby misclassification from class II–III obesity to class I obesity raised the HR in both categories.

DISCUSSION

We categorized BMI calculated from self-reported weight and height and BMI calculated from measured weight and height. We then compared the sex-specific hazard ratios for mortality obtained using self-reported BMI categories with the hazard ratios obtained using measured BMI categories in the same data set. We analyzed data separately for men and women, both for the full sample and for a restricted sample of healthy never-smokers. In the full sample, there was slight upward bias in hazard ratios at the highest BMI category. For men in the restricted sample, there was considerable upward bias in all BMI categories; for women in the restricted sample, there was downward bias in all BMI categories. Many studies have predicted that the effects of systematic error in self-reported BMI would be to increase hazard ratios at the

upper BMI levels because of the underreporting at the highest BMI levels (22, 23, 28, 53, 54). We observed this phenomenon, but we also observed that it was sometimes masked by the counter-effects of other types of misclassification.

Detailed examination of the impact of misclassifications between specific BMI categories revealed some effects that have not, to our knowledge, been previously appreciated. We found that the bias can be strongly affected by misclassification that affects the reference BMI category, because all hazard ratios are calculated relative to the referent. This is particularly important because of a tendency in large studies, often those using self-reported data, to argue in favor of using a “high-normal” reference category, such as BMI of 22.5–24.9, rather than the standard WHO normal weight BMI category of at least 18.5 and less 25 (32, 55, 56). However, when these categories are formed from self-reported BMI, using the narrower reference category increases the probability of misclassification. In the NHANES data, which has relatively accurate classification relative to other studies, the misclassification rate in the narrow reference category was close to 40%, and it may be higher in other studies.

Our results showed that the magnitude and even the direction of the bias in hazard ratios introduced by the use of self-reported data are not predictable, because the bias in hazard ratios arising from misclassification was related to the magnitude

Table 3. Effects on Mortality Hazard Ratios^a of Multiple Misclassifications That Affect the Specified Category of Body Mass Index, Using Data From National Health and Nutrition Examination Survey II (1976–1980), III (1988–1994), and Continuous (1999–2010), United States

BMI ^b Category Affected by Misclassification	BMI Category			
	Low (<22.5)	Overweight (25.0–29.9)	Class I Obesity (30.0–34.9)	Class II–III Obesity (≥35.0)
<i>Men</i>				
None (all measured data)	1.28	0.94	1.08	1.70
Low	1.30	0.95	1.09	1.72
Referent	1.30	0.94	1.09	1.72
Overweight	1.28	0.95	1.03	1.70
Class I obesity	1.28	0.96	1.06	1.83
Class II–III obesity	1.28	0.94	1.09	1.84
All (all self-reported data)	1.30	0.97	1.07	1.83
<i>Women</i>				
None (all measured data)	1.19	1.09	1.30	1.65
Low	1.18	1.11	1.32	1.67
Referent	1.17	1.12	1.31	1.66
Overweight	1.18	1.13	1.32	1.65
Class I obesity	1.19	1.11	1.36	1.66
Class II–III obesity	1.19	1.09	1.34	1.66
All (all self-reported data)	1.18	1.13	1.37	1.69

Abbreviation: BMI, body mass index.

^a Hazard ratios from Cox proportional hazards models, with age as the timeline and adjusted for survey, smoking status, racial/ethnic group, and alcohol consumption. The reference category was BMI of 22.5–24.9.

^b BMI was calculated as weight (kg)/height(m)². The combinations of individual misclassifications affecting each BMI category are shown in Web Table 1.

and direction of the true underlying hazard ratios using measured data. When the bias is a function of the measured hazard ratios, the bias in hazard ratios cannot be extrapolated to a different sample even if the 2 samples have the same degree of misclassification. Although some types of misreporting at upper levels of BMI increase hazard ratios, misreporting at lower levels of BMI may either further increase or diminish hazard ratios unpredictably. The high correlation coefficients and low mean differences often reported provide no evidence on the degree of bias.

An important finding of our study is that misclassification due to self-reported weight and height can reverse the direction of an association when the hazard ratio for overweight is below 1. A hazard ratio below 1 for overweight when measured data were used translated into a hazard ratio above 1 when self-reported data were used. Misclassification from the lower-risk overweight category into the reference category reduced the apparent risk in the reference category, thereby increasing the hazard ratios for overweight. The hazard ratio for overweight was further increased by misclassification from obesity to overweight. This may partially explain the tendency for studies using self-reported data to find higher hazard ratios for overweight than studies using measured data (20, 21). Shields et al. (26) similarly found a reversal of direction for overweight from 0.8 to 1.3 for the outcome of fair/poor health, and Dutton et al. (25),

using some of the same data, also found a number of examples of reversal of direction for overweight.

Two previous studies (47, 48), both also using the NHANES data, have looked at the relationship of self-reported versus measured BMI to mortality outcomes. Precise comparisons are limited because of differences in objectives and in analytical approaches. Keith et al. (47) compared statistical significance between categorizations of self-reported and measured BMI; Preston et al. (48) examined bias due to the combination of categorization and self-reported BMI data relative to continuous analyses with measured BMI data. Our findings are broadly consistent, but we provide a more detailed analysis of the effects of misclassification, particularly at lower BMI levels.

The bias in hazard ratios from using self-reported BMI was greater when the sample was restricted to healthy never-smokers. Keith et al. (47) also observed this in the NHANES data. The appearance of a greater impact in the restricted sample may be due in part to the higher hazard ratios for measured data in the restricted sample. There may be some additional effects of self-reporting of other variables that affect these results, because the information on smoking status and history of disease used to restrict the sample was also self-reported. Research suggests that inaccuracies in self-reported history of disease and inaccuracies in self-reported smoking status may be associated with various other characteristics (31, 54,

Table 4. Effects on Mortality Hazard Ratios^a of Multiple Misclassifications That Affect the Specified Category of Body Mass Index in the Restricted^b Sample of Healthy Never-Smokers, Using Data From National Health and Nutrition Examination Survey II (1976–1980), III (1988–1994), and Continuous (1999–2010), United States

BMI ^c Category Affected by Misclassification	Low (<22.5)	Overweight (25.0–29.9)	Class I Obesity (30.0–34.9)	Class II–III Obesity (≥35.0)
<i>Men</i>				
None (measured data)	0.87	0.85	1.13	1.86
Low	1.09	0.92	1.22	2.00
Referent	1.29	1.17	1.43	2.36
Overweight	1.02	1.14	1.17	2.16
Class I obesity	0.87	0.91	0.99	2.24
Class II–III obesity	0.87	0.86	1.11	2.26
All (self-reported data)	1.26	1.24	1.26	2.88
<i>Women</i>				
None (measured data)	1.37	1.21	1.59	1.94
Low	1.21	1.15	1.51	1.85
Referent	1.15	1.09	1.44	1.76
Overweight	1.30	1.18	1.56	1.84
Class I obesity	1.37	1.24	1.66	2.01
Class II–III obesity	1.37	1.21	1.60	2.00
All (self-reported data)	1.15	1.11	1.48	1.81

Abbreviation: BMI, body mass index.

^a Hazard ratios from Cox proportional hazards models, with age as the timeline and adjusted for survey, smoking status, racial/ethnic group, and alcohol consumption. The reference category was BMI of 22.5–24.9.

^b The restricted sample was limited to participants who reported never smoking, who reported no history of heart disease or cancer, and who were examined when younger than age 70 years.

^c BMI was calculated as weight (kg)/height(m)². The combinations of individual misclassifications affecting each BMI category are shown in Web Table 1.

57–59). Inaccuracies in the reporting of height and weight may also be associated with some features of health status (60). The factors that are associated with misreporting of height and weight, smoking status, or disease history may also be

associated with the outcome and thus generate additional residual confounding.

We selected a reference category of BMI of 22.5–24.9 often used in large studies, rather than the standard category

Table 5. Example of How a Sequence of Specified Category Misclassifications of Body Mass Index Changed the Direction of the Mortality Hazard Ratio^a for Overweight in the Restricted Sample^b of Men, Using Data From National Health and Nutrition Examination Survey II (1976–1980), III (1988–1994), and Continuous (1999–2010), United States

Number of BMI Misclassifications	Referent ^c to Low	Referent to Overweight	Overweight to Referent	Class I Obese to Overweight	HR for Overweight
None	No	No	No	No	0.85
1	No	Yes	No	No	0.97
2	No	Yes	Yes	No	1.08
3	No	Yes	Yes	Yes	1.17
4	Yes	Yes	Yes	Yes	1.24

Abbreviations: BMI, body mass index; HR, hazard ratio.

^a Hazard ratios from Cox proportional hazards models, with age as the timeline and adjusted for survey, smoking status, racial/ethnic group, and alcohol consumption. The reference category was BMI of 22.5–24.9.

^b The restricted sample was limited to participants who reported never smoking, who reported no history of heart disease or cancer, and who were examined when younger than age 70 years.

^c BMI was calculated as weight (kg)/height(m)². BMI classifications were as follows: <22.5 (low), 22.5–24.9 (reference), 25.0–29.9 (overweight), 30.0–34.9 (class I obesity), and ≥35.0 (class II–III obesity).

of normal BMI (18.5–24.9), which has been criticized on the basis that it is too broad (32, 55, 56). Discussions of the choice of reference category have not recognized that when BMI is self-reported a narrower reference category increases the probability of misclassification in the reference category, which may affect hazard ratios at higher BMI levels (55). The narrower the reference category, the more likely that the reference category is enriched by misclassification from the overweight category into the reference category. When self-reported data are used, the choice of narrower BMI categories will tend to increase possible bias from misclassification. A further implication is that, when self-reported data are used, approaches that involve restricting the low-BMI representation in the analytical sample, such as deleting respondents with BMI values below 20, might have an unexpected effect on the hazard ratios for higher BMI levels because of their effects on the reference category.

An important limitation of our analyses and other analyses that used data from NHANES (47, 48) is that the results presented here may underestimate the bias in other studies of BMI and mortality that have used self-reported BMI data. In NHANES, participants reported their own weight and height in a face-to-face interview knowing that their weight and height would also be measured within a few weeks. These reports may be more accurate than reported weight and height data collected through telephone or mail surveys or without an expectation of measurement (61, 62).

Several large studies of BMI and mortality have used self-reported weight and height data to estimate hazard ratios for BMI categories, sometimes in an attempt to identify the lowest-risk BMI category (2–4). However, the findings reported here suggest that self-reported BMI cannot reliably indicate the lowest-risk category because misclassification effects may misleadingly suggest the wrong category as the low category. In our data set, for men in the restricted sample, the lowest mortality was observed in the range of BMI 25–29.9 when measured data were used, but when self-reported data were used, the lowest mortality was observed in the reference category of 22.5–24.9. In a simulation study, Heavner and Burstyn (63) showed how the choice of varying cutoffs used to categorize an incorrectly measured exposure can affect odds ratios.

The results reported here confirm that, in general, the expected effect of underreporting at high BMI levels would be to bias hazard ratios upwards, but that effect can be augmented, counter-balanced, or even reversed by reporting errors that affect the reference category. The greater bias observed in the restricted sample of healthy never-smokers remains unexplained but may be related to the larger hazard ratios in the restricted sample and to some characteristics of misreporting of smoking status or health history.

ACKNOWLEDGMENTS

Author affiliations: Division of Health and Nutrition Examination Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland (Katherine M. Flegal, Brian K. Kit); Prevention Research Center, School of Medicine, Stanford University,

Palo Alto, California (Katherine M. Flegal); and Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland (Barry I. Graubard).

This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics.

We thank David Check for assistance with the figure.

The findings and conclusions in this report are those of the authors and not necessarily the official views of the Centers for Disease Control and Prevention or the National Cancer Institute.

Conflict of interest: none declared.

REFERENCES

1. Manson JE, Willett WC, Stampfer MJ, et al. Body weight and mortality among women. *N Engl J Med*. 1995;333(11):677–685.
2. Calle EE, Thun MJ, Petrelli JM, et al. Body-mass index and mortality in a prospective cohort of US adults. *N Engl J Med*. 1999;341(15):1097–1105.
3. Adams KF, Schatzkin A, Harris TB, et al. Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N Engl J Med*. 2006;355(8):763–778.
4. Berrington de Gonzalez A, Hartge P, Cerhan JR, et al. Body-mass index and mortality among 1.46 million white adults. *N Engl J Med*. 2010;363(23):2211–2219.
5. Finucane MM, Stevens GA, Cowan MJ, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet*. 2011;377(9765):557–567.
6. Connor Gorber S, Tremblay M, Moher D, et al. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev*. 2007;8(4):307–326.
7. Fillenbaum GG, Kuchibhatla MN, Whitson HE, et al. Accuracy of self-reported height and weight in a community-based sample of older African Americans and whites. *J Gerontol A Biol Sci Med Sci*. 2010;65(10):1123–1129.
8. Griebeler ML, Levis S, Beringer LM, et al. Self-reported versus measured height and weight in Hispanic and non-Hispanic menopausal women. *J Womens Health (Larchmt)*. 2011;20(4):599–604.
9. Gillum RF, Semplos CT. Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the Third National Health and Nutrition Examination Survey. *Nutr J*. 2005;4:27.
10. Spencer EA, Appleby PN, Davey GK, et al. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr*. 2002;5(4):561–565.
11. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, et al. Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *J Epidemiol Community Health*. 2000;54(2):143–148.
12. Nyholm M, Gullberg B, Merlo J, et al. The validity of obesity based on self-reported weight and height: implications for population studies. *Obesity (Silver Spring)*. 2007;15(1):197–208.

13. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? *Public Health*. 1995;109(4):275–284.
14. Shields M, Connor Gorber S, Tremblay MS. Estimates of obesity based on self-report versus direct measures. *Health Rep*. 2008;19(2):61–76.
15. Visscher TL, Viet AL, Kroesbergen IH, et al. Underreporting of BMI in adults and its effect on obesity prevalence estimations in the period 1998 to 2001. *Obesity (Silver Spring)*. 2006;14(11):2054–2063.
16. Yun S, Zhu BP, Black W, et al. A comparison of national estimates of obesity prevalence from the behavioral risk factor surveillance system and the National Health and Nutrition Examination Survey. *Int J Obes (Lond)*. 2006;30(1):164–170.
17. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol*. 1986;123(4):736–751.
18. Manson JE, Bassuk SS, Hu FB, et al. Estimating the number of deaths due to obesity: can the divergent findings be reconciled? *J Womens Health (Larchmt)*. 2007;16(2):168–176.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
20. Flegal KM, Kit BK, Orpana H, et al. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA*. 2013;309(1):71–82.
21. Flegal KM, Kit BK, Graubard BI. Body mass index categories in observational studies of weight and risk of death. *Am J Epidemiol*. 2014;180(3):288–296.
22. Chiolero A, Peytremann-Bridevaux I, Paccaud F. Associations between obesity and health conditions may be overestimated if self-reported body mass index is used. *Obes Rev*. 2007;8(4):373–374.
23. Faeh D, Roh L, Paccaud F, et al. Mortality risk of obesity and underweight is overestimated with self-reported body mass index. *Epidemiology*. 2014;25(1):156–158.
24. James WPT, Jackson-Leach R, Ni Mhurchu C, et al. 8. Overweight and obesity (high body mass index). In: Ezzati M, Lopez AD, Rodgers A, et al., eds. *Comparative Quantification of Health Risks*. Vol. 1. Geneva, Switzerland: World Health Organization; 2004:497–596.
25. Dutton DJ, McLaren L. The usefulness of “corrected” body mass index vs. self-reported body mass index: comparing the population distributions, sensitivity, specificity, and predictive utility of three correction equations using Canadian population-based data. *BMC Public Health*. 2014;14:430.
26. Shields M, Gorber SC, Tremblay MS. Effects of measurement on obesity and morbidity. *Health Rep*. 2008;19(2):77–84.
27. Yannakoulia M, Panagiotakos DB, Pitsavos C, et al. Correlates of BMI misreporting among apparently healthy individuals: the ATTICA study. *Obesity (Silver Spring)*. 2006;14(5):894–901.
28. Boffetta P, McLaughlin JK, la Vecchia C, et al. “Environment” in cancer causation and etiological fraction: limitations and ambiguities. *Carcinogenesis*. 2007;28(5):913–915.
29. Heitmann BL. Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. Implications for diet disease relationships. *Int J Epidemiol*. 1996;25(1):222–225.
30. Klatsky AL, Udaltsova N, Li Y, et al. Moderate alcohol intake and cancer: the role of underreporting. *Cancer Causes Control*. 2014;25(6):693–699.
31. Fisher MA, Taylor GW, Shelton BJ, et al. Sociodemographic characteristics and diabetes predict invalid self-reported non-smoking in a population-based study of US adults. *BMC Public Health*. 2007;7:33.
32. Lewis CE, McTigue KM, Burke LE, et al. Mortality, health outcomes, and body mass index in the overweight range: a science advisory from the American Heart Association. *Circulation*. 2009;119(25):3263–3271.
33. Batsis JA, Sahakyan KR, Rodriguez-Escudero JP, et al. Normal weight obesity and mortality in United States subjects ≥ 60 years of age (from the Third National Health and Nutrition Examination Survey). *Am J Cardiol*. 2013;112(10):1592–1598.
34. Batsis JA, Singh S, Lopez-Jimenez F. Anthropometric measurements and survival in older Americans: results from the Third National Health and Nutrition Examination Survey. *J Nutr Health Aging*. 2014;18(2):123–130.
35. Flegal KM, Graubard BI. Estimates of excess deaths associated with body mass index and other anthropometric variables. *Am J Clin Nutr*. 2009;89(4):1213–1219.
36. Flegal KM, Graubard BI, Williamson DF, et al. Excess deaths associated with underweight, overweight, and obesity. *JAMA*. 2005;293(15):1861–1867.
37. Flegal KM, Graubard BI, Williamson DF, et al. Cause-specific excess deaths associated with underweight, overweight, and obesity. *JAMA*. 2007;298(17):2028–2037.
38. Flegal KM, Graubard BI, Williamson DF, et al. Impact of smoking and preexisting illness on estimates of the fractions of deaths associated with underweight, overweight, and obesity in the US population. *Am J Epidemiol*. 2007;166(8):975–982.
39. Kuk JL, Ardern CI. Are metabolically normal but obese individuals at lower risk for all-cause mortality? *Diabetes Care*. 2009;32(12):2297–2299.
40. Kuk JL, Ardern CI. Influence of age on the association between various measures of obesity and all-cause mortality. *J Am Geriatr Soc*. 2009;57(11):2077–2084.
41. Reis JP, Araneta MR, Wingard DL, et al. Overall obesity and abdominal adiposity as predictors of mortality in US white and black adults. *Ann Epidemiol*. 2009;19(2):134–142.
42. Reis JP, Macera CA, Araneta MR, et al. Comparison of overall obesity and body fat distribution in predicting risk of mortality. *Obesity (Silver Spring)*. 2009;17(6):1232–1239.
43. Sharma S, Batsis JA, Coutinho T, et al. Normal-weight central obesity and mortality risk in older adults with coronary artery disease. *Mayo Clin Proc*. 2016;91(3):343–351.
44. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988–94. Series 1: programs and collection procedures. *Vital Health Stat 1*. 1994;(32):1–407.
45. McDowell A, Engel A, Massey JT, et al. Plan and operation of the Second National Health and Nutrition Examination Survey, 1976–1980. *Vital Health Stat 1*. 1981;(15):1–144.
46. Zipf G, Chiappa M, Porter KS, et al. National Health and Nutrition Examination Survey: plan and operations, 1999–2010. *Vital Health Stat 1*. 2013;(56):1–37.
47. Keith SW, Fontaine KR, Pajewski NM, et al. Use of self-reported height and weight biases the body mass index-mortality association. *Int J Obes (Lond)*. 2011;35(3):401–408.
48. Preston SH, Fishman E, Stokes A. Effects of categorization and self-report bias on estimates of the association between obesity and mortality. *Ann Epidemiol*. 2015;25(12):907–911.e2.
49. Feinstein AR, Sosin DA, Wells CK. The Will Rogers phenomenon: improved technologic diagnosis and stage migration as a source of nontherapeutic improvement in cancer prognosis. *Trans Assoc Am Physicians*. 1984;97:19–24.
50. Sormani MP. The Will Rogers phenomenon: the effect of different diagnostic criteria. *J Neurol Sci*. 2009;287(suppl 1):S46–S49.

51. Thompson IM, Canby-Hagino E, Lucia MS. Stage migration and grade inflation in prostate cancer: Will Rogers meets Garrison Keillor. *J Natl Cancer Inst.* 2005;97(17):1236–1237.
52. Young MJ, Lenhart J, Wasser TE, et al. Evidence for the Will Rogers phenomenon in migration of employees to managed care plans. *J Gen Intern Med.* 1999;14(9):564–566.
53. Bogers RP, Bemelmans WJ, Hoogenveen RT, et al. Association of overweight with increased risk of coronary heart disease partly independent of blood pressure and cholesterol levels: a meta-analysis of 21 cohort studies including more than 300 000 persons. *Arch Intern Med.* 2007;167(16):1720–1728.
54. Park JY, Mitrou PN, Keogh RH, et al. Effects of body size and sociodemographic characteristics on differences between self-reported and measured anthropometric data in middle-aged men and women: the EPIC-Norfolk study. *Eur J Clin Nutr.* 2011;65(3):357–367.
55. Joshy G, Korda RJ, Bauman A, et al. Investigation of methodological factors potentially underlying the apparently paradoxical findings on body mass index and all-cause mortality. *PLoS One.* 2014;9(2):e88641.
56. Tobias DK, Hu FB. Does being overweight really reduce mortality? *Obesity (Silver Spring).* 2013;21(9):1746–1749.
57. Muggah E, Graves E, Bennett C, et al. Ascertainment of chronic diseases using population health data: a comparison of health administrative data and patient self-report. *BMC Public Health.* 2013;13:16.
58. Navin Cristina TJ, Stewart Williams JA, Parkinson L, et al. Identification of diabetes, heart disease, hypertension and stroke in mid- and older-aged women: comparing self-report and administrative hospital data records. *Geriatr Gerontol Int.* 2016;16(1):95–102.
59. Suadicani P, Hein HO, Gyntelberg F. Mortality and morbidity of potentially misclassified smokers. *Int J Epidemiol.* 1997;26(2):321–327.
60. Katsnelson MJ, Peterlin BL, Rosso AL, et al. Self-reported vs measured body mass indices in migraineurs. *Headache.* 2009;49(5):663–668.
61. Ezzati M, Martin H, Skjold S, et al. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *J R Soc Med.* 2006;99(5):250–257.
62. Scribani M, Shelton J, Chapel D, et al. Comparison of bias resulting from two methods of self-reporting height and weight: a validation study. *JRSM Open.* 2014;5(6):2042533313514048.
63. Heavner K, Burstyn I. A simulation study of categorizing continuous exposure variables measured with error in autism research: small changes with large effects. *Int J Environ Res Public Health.* 2015;12(8):10198–10234.