Accuracy and reliability of self-reported weight and height in the Sister Study

Cynthia J Lin, Lisa A DeRoo, Sara R Jacobs and Dale P Sandler*
Epidemiology Branch, National Institute of Environmental Health Sciences, PO Box 12233, MD A3-05, Research Triangle Park, NC 27709, USA

Abstract

Objective: To assess the accuracy and reliability of self-reported weight and height and identify the factors associated with reporting accuracy.
Design: Analysis of self-reported and measured weight and height from participants in the Sister Study (2003–2009), a nationwide cohort of 50 884 women aged 35–74 years in the USA with a sister with breast cancer.
Setting: Weight and height were reported via computer-assisted telephone interview (CATI) and self-administered questionnaires, and measured by examiners.
Subjects: Early enrolees in the Sister Study. There were 18 639 women available for the accuracy analyses and 13 316 for the reliability analyses.
Results: Using weighted kappa statistics, comparisons were made between CATI responses and examiner measures to assess accuracy and CATI and questionnaire responses to assess reliability. Polytomous logistic regression evaluated factors associated with over- or under-reporting. Compared with measured values, agreement was 96 % for reported height (± 1 inch ($\pm 2\cdot 5$ cm); weighted $\kappa = 0\cdot 84$) and 67 % for weight (± 3 lb ($\pm 1\cdot 36$ kg); weighted $\kappa = 0\cdot 92$). Obese women (BMI ≥ 30 kg/m^2) were more likely than normal-weight women to under-report weight by ≥ 5 % and underweight women (BMI $< 18\cdot 5$ kg/m^2) were more likely to over-report. Among normal-weight and overweight women ($18\cdot 5$ kg/m$^2 \leq$ BMI < 30 kg/m^2), weight cycling and lifetime weight difference ≥ 50 lb ($\geq 22\cdot 68$ kg) were associated with over-reporting.
Conclusions: US women in the Sister Study were reasonably reliable and accurate in reporting weight and height. Women with normal-range BMI reported most accurately. Overweight and obese women and those with weight fluctuations were less accurate, but even among obese women, few under-reported their weight by > 10 %.

Many studies have found an association between high or low BMI and risk of adverse health outcomes using self-reported data on weight and height. With an increasing prevalence of overweight and obesity in the USA[1], the effect of anthropometric characteristics on reporting accuracy is a concern. Studies have examined the accuracy of self-reported *v.* directly measured height and weight but findings have varied and many studies were small or otherwise limited[2,3]. In a meta-analysis of weight reporting in thirty-four studies, only eighteen were from the USA, sample sizes varied from eighteen to 9000, ages varied from 12 to 84 years, and measurement protocols differed or were not described[2]. While many studies suggest that women tend to under-report their weight, less is known about the factors associated with reporting accuracy.

Current weight has been shown to influence weight reporting accuracy. The overweight and obese tend to under-report their weight and the underweight tend to over-report[4,5]. Studies of selected populations, including adult women in the USA, have also suggested that age and race contribute to reporting bias[6–8].

The impact of weight fluctuation and weight cycling on weight reporting accuracy has not been thoroughly examined in the existing literature. Weight cycling is not uncommon. Among Finnish women, the prevalence of weight cycling (defined as losing and then regaining ≥ 5 kg) was reported to be 29 %[9]. Strohacker *et al.* estimated that 38 % of US women weight cycle at least once in their lifetime[10], and 20 % of women in the Nurses' Health Study reported at least three weight cycling episodes (defined as losing and then regaining ≥ 10 lb ($\geq 4\cdot 54$ kg))[11]. Among obese bariatric surgery candidates, frequent weight cycling was associated with greater reporting accuracy, suggesting that frequent weight cycling might increase attentiveness to weight, leading to heightened accuracy in reporting[12]. Weight cycling and fluctuation and weight

Corresponding author: Email sandler@niehs.nih.gov

reporting accuracy have not yet been examined in a large sample of the general population.

A tendency to over-report height has been observed, particularly among people who are older, shorter and/or overweight[8], but under-reporting has been observed in higher income categories for certain age groups[13]. Fewer studies have assessed reliability of self-reported measures and results were inconsistent[14,15].

The present study assessed the accuracy and reliability of self-reported weight and height in a large cohort of US women and identified characteristics associated with reporting accuracy. We compared self-reported height and weight with examiner-measured values, and separately compared two self-reports obtained using different approaches, allowing us to consider design features affecting data quality.

Methods

Data collection and study population

We used data from the Sister Study, a nationwide volunteer cohort of 50 884 US (including Puerto Rico) women aged 35–74 years with a sister with breast cancer; enrolment occurred from September 2003 to March 2009. The present analysis examines early enrolees who completed baseline activities by 21 September 2007 (*n* 31 409). To avoid errors influenced by eating disorders[16–18], participants who reported ever having anorexia or bulimia were excluded (*n* 1066). Pregnant women delayed baseline activities until at least three months after the end of pregnancy.

Study participants reported weight (pounds) and height (feet and inches) in a computer-assisted telephone interview (CATI) and separately on a self-administered scannable diet questionnaire. During a home visit, trained examiners used digital self-calibrating scales to measure weight and metal tape measures to measure height. The order of completing the CATI, questionnaire and home visit varied; self-reports could be completed before or after the home visit. All measurements were taken three times without shoes. Measurements were rounded to the nearest whole pound for weight and quarter of an inch for height. Other variables examined from the baseline CATI were weight cycling (frequency of losing and then gaining ≥ 20 lb ($\geq 9{\cdot}07$ kg)), lowest weight since age 20 years, heaviest non-pregnant/breast-feeding weight, age, race, education level, perceived health status, marital status, household income, smoking, alcohol, physical activity, gravidity, regular multivitamin intake, recency of last medical examination, history of depression and use of antidepressant medications.

BMI was categorized using the Centers for Disease Control and Prevention definitions[19]. Lifetime weight difference was calculated by subtracting lowest weight since age 20 years from heaviest non-pregnant/breast-feeding weight.

All statistical analyses were performed using the STATA/IC for Windows statistical software package version $10{\cdot}1$ (StataCorp LP, College Station, TX, USA).

Accuracy of telephone interview

To assess the accuracy of self-reported weight and height, we first compared CATI-reported values with examiner measures among women who completed the CATI within 30 d of the home visit (*n* 18 639). The primary source of Sister Study data is the telephone interview, which had less missing data and fewer structural errors (see below) for height and weight. For this analysis, examiner measures were treated as the true value. Percentage agreement and weighted kappa statistics were calculated for each variable of interest. Kappa statistics were weighted according to a standard weight in STATA to account for the degree of disagreement. Polytomous logistic regression was used to calculate odds ratios and 95 % confidence intervals for reporting accuracy by age, race, education level, perceived health status, marital status and measured BMI.

To be consistent with the existing literature, we first examined the absolute difference between self-reported and measured weight. Differences between measured and self-reported weight were categorized as under-reporting by ≥ 7 lb ($\geq 3{\cdot}18$ kg), under-reporting by 4–6 lb ($1{\cdot}81$–$2{\cdot}72$ kg), reporting within 3 lb ($1{\cdot}36$ kg) and over-reporting by ≥ 4 lb ($\geq 1{\cdot}81$ kg). Because the relative impact of a specific weight difference will be greater in smaller than larger women, we also calculated the percentage of weight misreported; self-reports that differed by less than 5 % from measured weights were the referent category. Polytomous logistic regression models explored the effects of measured BMI, weight cycling, lifetime weight difference and current antidepressant use on under- and over-reporting, adjusting for age, race, education, perceived health status and marital status as potential confounders. Models examining weight cycling, lifetime weight difference or current antidepressant use also adjusted for measured BMI. Differences between measured and self-reported height were categorized as under-reporting by ≥ 1 inch ($\geq 2{\cdot}5$ cm), reporting within 1 inch and over-reporting by ≥ 1 inch.

To determine the effect of misreporting on BMI categories, we compared categories calculated from CATI-reported data with categories based on examiner-measured data using percentage agreement and weighted kappa statistics for all women and stratified by categories of age, race, education level, perceived health status and marital status. We also determined the sensitivity and specificity of self-reported overweight/obese classification relative to examiner-measured data. To further explore the potential for bias in BMI we stratified on measured BMI and examined the percentage of CATI-determined BMI values that over- or underestimated BMI calculated from examiner-measured values.
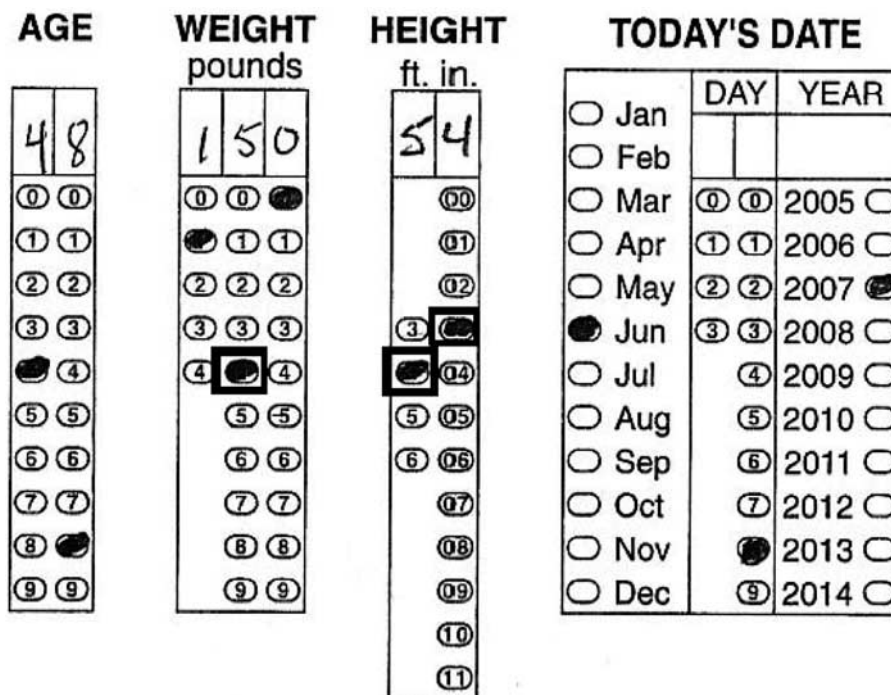
Fig. 1 Example of frameshift errors on the self-administered diet questionnaire

We carried out additional analyses stratifying by or adjusting for which measure came first, the home visit or CATI.

Accuracy of self-completed questionnaire

Using data from the subset of women with CATI and questionnaire completed within 30 d of the home visit (*n* 13 985), we carried out similar analyses to assess the accuracy of weight and height reported in the self-completed questionnaire compared with examiner-measured data. We then compared the accuracy of the two self-report measures by calculating ratios of odds ratios from models assessing reporting by CATI or questionnaire *v.* measured data. An analysis including all women (*n* 21 935) completing the diet questionnaire within 30 d of the home visit had similar results and is not shown.

Reliability

Reliability of self-reported weight and height was assessed using percentage agreement and weighted kappa statistics to compare self-reported data from the CATI and diet questionnaires. Analyses were limited to women who completed the CATI within 30 d of submitting their questionnaire (*n* 13 316) and had non-missing questionnaire data for weight (*n* 11 585) and height (*n* 11 885). Similar to the accuracy analysis, we stratified and adjusted analyses by reporting order with respect to each other and with respect to the examiner measurement.

Correcting structural errors

Prior to analyses, we identified and corrected several problems inherent to the reporting method. Both random and systematic errors occurred with the self-administered diet questionnaire. About 1 % of respondents appeared to make frameshift bubbling errors for weight and/or height by mistaking the bubbles in one or more columns as starting at 1 instead of 0. Figure 1 shows a frameshift error in which the respondent filled in the wrong value for weight in the tens place and the wrong values for height in the feet and inches columns. Frameshift errors occurred frequently in the hundreds place of weight, which were detected when an unreasonable weight (<100 lb) was marked (e.g. 34 lb instead of 134 lb). We corrected obvious frameshift errors (0·7 % of weight values and 0·1 % of height values) when questionnaire values differed from both the CATI and examiner reports by >60 lb (>27·22 kg) or >11 inches (>27·9 cm).

Some errors were related to the choice of unit. In the diet questionnaires, a small percentage of respondents appeared to report height in total inches rather than feet and inches as instructed. For example, instead of 5 feet 4 inches, a respondent marked the total inch equivalent (64 inches) which was then mistakenly interpreted as 6 feet 4 inches. We corrected these unit errors in about 0·8 % of all responses by checking suspiciously high reports and confirming corrections with CATI and examiner reports. Although these errors occurred for units (inches, pounds) used in the USA, similar errors could occur for those used in other countries (e.g. metres, kilograms).

There were considerable missing values for weight (13%), height (11%) or both (8%) in the self-administered diet questionnaires. Non-response did not substantially vary by age or BMI category. Missing weight and height were uncommon in the CATI (<1%).

There seemed to be a tendency to round to 0 or 5 when reporting weight in the CATI (59%) and questionnaires (52%), whereas an end digit of 0 or 5 occurred in 27% of examiner measures. We did not correct for this apparent rounding.

We detected infrequent random reporting errors for all modes of reporting. In self-administered questionnaires, random bubbling errors such as pencil smudges were sensitive to the questionnaire scanner. For the CATI, there were occasional data-entry errors by interviewers and for examiners, some inconsistencies following measurement protocols. We corrected CATI values if they greatly differed from both examiner and questionnaire values ($\geq$100 lb ($\geq$45·36 kg) for weight; $\geq$11 inches ($\geq$27·9 cm) for height).

Results

Participant characteristics

Participants were predominantly white (93%), aged 45–64 years (70%), college-educated (>50%), and married or living as married (77%; Table 1). Over half (58%) were overweight or obese; 78% perceived themselves as being in very good or excellent health.

Accuracy of telephone interview weight

Measured and self-reported (CATI) weight were highly correlated (correlation coefficient, $r = 0.99$). Overall, women under-reported their weight by an average of 1·6 lb (0·73 kg). The mean absolute difference between measured and CATI weight was 3·3 (SD 4·1; range 0–50) lb (1·50 (SD 1·86; range 0–22·68) kg). Mean self-reported weight was 160·2 (SD 35·5; range 82–402) lb (72·67 (SD 16·10; range 37·20–182·34) kg); mean examiner-measured weight was 161·8 (SD 36·4; range 80–425) lb (73·39 (SD 16·51; range 36·29–192·78) kg). The average absolute time between the CATI and examiner home visit was 12·6 (SD 8·7) d.

Overall, 66·5% of women reported their weight within 3 lb (1·36 kg) of measured values (Table 1) with overall weighted $\kappa = 0.92$. Agreement within 3 lb increased with age and perceived health status and was greater for women who were married, had a college degree and had normal measured BMI. Agreement was lower for black women, obese women, women who weight cycled $\geq$3 times and women who completed the CATI before the physical exam.

Under-reporting weight

The crude odds ratio for under-reporting by $\geq$7 lb ($\geq$3·18 kg) decreased with increasing age: OR = 0·84 (95% CI 0·75, 0·94) for women aged 55–64 years and OR = 0·62 (95% CI 0·54, 0·73) for women over 65 years,

compared with those aged 45–54 years (Table 2). Compared with non-Hispanic whites, blacks had a higher odds of under-reporting weight (OR = 1·26; 95% CI 1·00, 1·59 for under-reporting by 4–6 lb pounds (1·81–2·72 kg) and OR = 1·72; 95% CI 1·36, 2·17 for under-reporting by $\geq$7 lb). Never married (OR = 1·41; 95% CI 1·16, 1·72) and widowed/divorced/separated women (OR = 1·25; 95% CI 1·11, 1·40) had an increased odds of under-reporting weight by $\geq$7 lb than married women. The odds ratio for under-reporting by $\geq$7 lb increased from 3·82 (95% CI 3·29, 4·43) for overweight women to 8·92 (95% CI 7·74, 10·29) for obese relative to normal-weight women. Associations remained after adjusting for age, race and education (Table 2); further adjustment for perceived health and marital statuses did not substantially change estimates. Results from analyses stratified by reporting order (CATI before or after physical exam) were similar.

The effect of weight cycling differed by BMI status, affecting mainly reporting accuracy among underweight and normal-weight women (Table 3).

About 8% of all women (*n* 1439) under-reported weight by $\geq$5%. Compared with normal-weight women, in adjusted analyses, the odds of under-reporting weight by $\geq$5% was higher among overweight (OR = 2·38; 95% CI 2·05, 2·77) and obese women (OR = 4·10; 95% CI 3·54, 4·76; Fig. 2). A lifetime weight difference of 25–49 lb (11·34–22·23 kg) was also associated with under-reporting (OR = 1·35; 95% CI 1·11, 1·65; Fig. 3). Stratifying by BMI, overweight and obese women with a lifetime weight difference of $\geq$50 lb ($\geq$22·68 kg) had a decreased odds of under-reporting weight by $\geq$5% compared with those with a smaller weight difference (overweight OR = 0·65; 95% CI 0·54, 0·78; obese OR = 0·52; 95% CI 0·39, 0·70). Conversely, underweight and normal-weight women who weight cycled at least once had an increased odds of under-reporting weight compared with those who never weight cycled (OR = 1·35; 95% CI 1·02, 1·78).

Over-reporting weight

Only 2% (*n* 465) of all women over-reported weight by $\geq$5%. In adjusted analyses, the most important factor associated with over-reporting weight by $\geq$5% was being underweight: OR = 5·30 (95% CI 3·67, 7·66; Fig. 2). Weight cycling and increasing lifetime weight difference were also associated with over-reporting weight (Fig. 3).

After excluding currently underweight and obese women, the increased odds of over-reporting by $\geq$5% remained for those having a lifetime difference of $\geq$75 lb ($\geq$34·02 kg; OR = 2·89; 95% CI 1·76, 4·75; data not shown). However, the increased odds of over-reporting among women with $\geq$3 episodes of weight cycling was no longer significant (OR = 1·30; 95% CI 0·89, 1·90; data not shown). After stratifying by BMI, lifetime weight difference $\geq$50 lb ($\geq$22·68 kg) was associated with over-reporting among currently normal-weight (OR = 1·73; 95% CI 1·22, 2·46) and overweight women (OR 1·58; 95% CI 1·05, 2·38).

Table 1 Characteristics of participants by weight reporting accuracy: US women (n 18 639) aged 35–74 years, the Sister Study (2003–2009)

Characteristic	All		Under-report by ≥7 lb (≥3·18 kg) (n 1996)	Under-report by 4–6 lb (1·81–2·72 kg) (n 2679)	Report within 3 lb (1·36 kg) (n 12 394)	Over-report by ≥4 lb (≥1·81 kg) (n 1570)	Weighted κ
	n	%	%	%	%	%	
All women	18 639	100	10·7	14·4	66·5	8·4	0·92
Age							
<45 years	2471	13·3	12·0	15·1	64·5	8·5	0·91
45–54 years	6305	33·8	12·1	14·3	65·5	8·2	0·92
55–64 years	6571	35·3	10·3	14·7	66·7	8·3	0·92
≥65 years	3292	17·7	8·0	13·5	69·4	9·1	0·92
Race							
White, non-Hispanic	17 250	92·5	10·5	14·3	66·7	8·4	0·92
Black, Hispanic or non-Hispanic	594	3·2	16·2	16·2	59·6	8·1	0·90
Hispanic	332	1·8	11·1	12·7	67·5	8·7	0·92
Other	461	2·5	10·2	15·0	66·6	8·2	0·92
Education							
Less than bachelors	8489	45·5	10·8	14·5	65·5	9·2	0·92
Bachelors	5359	28·8	10·3	14·2	67·8	7·8	0·92
Masters/doctorate	4788	25·7	11·0	14·3	66·9	7·7	0·91
Perceived health status							
Very good/excellent	14 490	77·7	10·2	14·6	67·4	7·7	0·91
Good	3212	17·2	12·4	13·6	63·2	10·8	0·92
Fair/poor	935	5·0	12·5	13·3	63·1	11·1	0·93
Marital status							
Never married	940	5·0	13·8	10·7	64·9	10·5	0·93
Legally married/living as married	14 343	77·0	10·1	14·7	66·9	8·2	0·91
Widowed/divorced/separated	3353	18·0	12·3	13·9	65·1	8·8	0·92
BMI							
Underweight (<18·5 kg/m^2)	241	1·3	0·0	5·0	71·8	23·2	0·76
Normal weight (18·5–24·9 kg/m^2)	7602	40·8	3·4	11·2	75·9	9·5	0·85
Overweight (25·0–29·9 kg/m^2)	5859	31·4	11·3	17·0	64·9	6·8	0·82
Obese (≥30·0 kg/m^2)	4937	26·5	21·8	16·6	53·7	8·0	0·86
Current antidepressant use							
No	16 476	88·4	10·5	14·4	66·8	8·3	0·91
Yes	2047	11·0	13·0	13·8	63·5	9·6	0·92
Weight cycling							
Never	10 338	55·5	7·3	13·0	71·4	8·3	0·90
1–2 time(s)	4687	25·1	12·0	16·5	63·4	8·1	0·90
≥3 times	3614	19·4	18·8	15·6	56·5	9·1	0·90
Reporting order							
CATI first, then physical exam	6912	37·1	14·9	15·8	57·7	11·6	0·90
Physical exam first, then CATI	11 052	59·3	8·5	13·8	70·9	6·7	0·93
Same day CATI and physical exam	675	3·6	3·7	8·4	84·3	3·6	0·96

CATI, computer-assisted telephone interview.

Table 2 Association of participant characteristics with weight reporting accuracy – CATI-reported *v.* examiner-measured weight: US women (*n* 18 639) aged 35–74 years, the Sister Study (2003–2009)

Characteristic	Under-report by ≥7 lb (≥3·18 kg)				Under-report by 4–6 lb (1·81–2·72 kg)				Over-report by ≥4 lb (≥1·81 kg)			
	OR	95 % CI	aOR	95 % CI	OR	95 % CI	aOR	95 % CI	OR	95 % CI	aOR	95 % CI
Age												
<45 years	1·01	0·87, 1·17	1·01	0·87, 1·17	1·07	0·94, 1·23	1·07	0·94, 1·22	1·05	0·89, 1·25	1·06	0·89, 1·26
45–54 years	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
55–64 years	0·84	0·75, 0·94	0·84	0·75, 0·94	1·01	0·91, 1·12	1·01	0·91, 1·12	1·00	0·88, 1·14	1·01	0·89, 1·14
≥65 years	0·62	0·54, 0·73	0·62	0·54, 0·72	0·89	0·79, 1·01	0·89	0·79, 1·01	1·05	0·90, 1·22	1·04	0·89, 1.21
Race												
White, non-Hispanic	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
Black, Hispanic or non-Hispanic	1·72	1·36, 2·17	1·69	1·34, 2·13	1·26	1·00, 1·59	1·26	1·00, 1·58	1·07	0·79, 1·46	1·08	0·80, 1·47
Hispanic	1·05	0·74, 1·49	0·97	0·68, 1·38	0·87	0·63, 1·22	0·85	0·61, 1·19	1·02	0·69, 1·51	1·00	0·68, 1·48
Other	0·97	0·71, 1·32	0·95	0·69, 1·30	1·05	0·80, 1·36	1·04	0·80, 1·35	0·98	0·70, 1·38	0·96	0·69, 1·36
Education												
Less than bachelors	1·09	0·97, 1·22	1·13	1·01, 1·27	1·06	0·96, 1·17	1·07	0·97, 1·19	1·22	1·08, 1·38	1·23	1·08, 1.39
Bachelors	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
Masters/doctorate	1·09	0·96, 1·24	1·12	0·98, 1·27	1·03	0·91, 1·15	1·03	0·92, 1·16	1·01	0·87, 1·17	1·01	0·87, 1·17
Perceived health status												
Very good/excellent	0·78	0·69, 0·88	0·76	0·67, 0·86	1·01	0·90, 1·13	1·01	0·90, 1·13	0·67	0·59, 0·76	0·68	0·59, 0.77
Good	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
Fair/poor	1·01	0·81, 1·27	1·02	0·82, 1·28	0·98	0·79, 1·22	0·98	0·79, 1·22	1·03	0·81, 1·30	1·03	0·81, 1.3
Marital status												
Never married	1·41	1·16, 1·72	1·36	1·11, 1·66	0·75	0·61, 0·93	0·75	0·6, 0·93	1·32	1·06, 1·65	1·38	1·10, 1.72
Legally married/living as married	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
Widowed/divorced/separated	1·25	1·11, 1·40	1·34	1·19, 1·51	0·97	0·87, 1·08	0·99	0·89, 1·11	1·10	0·96, 1·26	1·10	0·96, 1·27
BMI												
Underweight (<18·5 kg/m²)	–	–	–	–	0·47	0·26, 0·85	0·47	0·26, 0·84	2·59	1·90, 3·54	2·60	1·91, 3.55
Normal weight (18·5–24·9 kg/m²)	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
Overweight (25·0–29·9 kg/m²)	3·82	3·29, 4·43	4·06	3·50, 4·72	1·77	1·60, 1·95	1·81	1·63, 2·00	0·84	0·74, 0·96	0·83	0·73, 0·94
Obese (≥30·0 kg/m²)	8·92	7·74, 10·29	9·59	8·30, 11·08	2·09	1·88, 2·33	2·14	1·92, 2·38	1·19	1·04, 1·36	1·16	1·01, 1·32
Current antidepressant use												
No	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
Yes	1·31	1·14, 1·51	1·30	1·13, 1·49	1·01	0·88, 1·16	1·00	0·88, 1·15	1·22	1·04, 1·44	1·23	1·04, 1·44
Reporting order												
CATI first, then physical exam	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.	1·00	Ref.
Physical exam first, then CATI	0·47	0·43, 0·52	0·47	0·43, 0·52	0·71	0·65, 0·78	0·71	0·65, 0·78	0·48	0·43, 0·53	0·47	0·43, 0·53
Same day CATI and physical exam	0·17	0·11, 0·26	0·17	0·11, 0·26	0·37	0·28, 0·48	0·36	0·28, 0·48	0·21	0·14, 0·32	0·21	0·14, 0·32

CATI, computer-assisted telephone interview; OR, crude odds ratio; aOR adjusted odds ratio (adjusted for age, race and education); Ref., referent group.
Participants reporting within 3 lb (1·36 kg) is the referent group.

Table 3 Association of weight cycling with weight reporting accuracy, stratified by BMI: US women (n 18 639) aged 35–74 years, the Sister Study (2003–2009)

BMI	Weight cycling	n	Under-report by ≥7 lb (≥3·18 kg)		Under-report by 4–6 lb (1·81–2·72 kg)		Over-report by ≥4 lb (≥1·81 kg)	
			aOR	95 % CI	aOR	95 % CI	aOR	95 % CI
Underweight (<18·5 kg/m²)	Never	223	1·00	Ref.	1·00	Ref.	1·00	Ref.
	1–2 time(s)	17	–	–	5·79	1·02, 32·98	5·34	1·80, 15·81
	≥3 times	1	–	–	–	–	–	–
Normal weight (18·5–24·9 kg/m²)	Never	6194	1·00	Ref.	1·00	Ref.	1·00	Ref.
	1–2 time(s)	1116	1·44	1·04, 2·00	1·22	1·00, 1·48	1·31	1·06, 1·61
	≥3 times	292	1·86	1·09, 3·16	1·23	0·86, 1·77	1·09	0·73, 1·64
Overweight (25·0–29·9 kg/m²)	Never	2989	1·00	Ref.	1·00	Ref.	1·00	Ref.
	1–2 time(s)	1879	0·87	0·72, 1·05	1·00	0·86, 1·17	1·06	0·84, 1·35
	≥3 times	991	1·06	0·84, 1·33	1·02	0·84, 1·24	1·34	1·02, 1·76
Obese (≥30·0 kg/m²)	Never	932	1·00	Ref.	1·00	Ref.	1·00	Ref.
	1–2 time(s)	1675	0·82	0·67, 1·01	1·12	0·89, 1·40	1·09	0·78, 1·52
	≥3 times	2330	1·10	0·91, 1·33	1·02	0·82, 1·26	1·60	1·18, 2·18

aOR, adjusted odds ratio (adjusted for age, race and education); Ref., referent group.
Participants reporting within 3 lb (1·36 kg) is the referent group.

Factors not associated with weight reporting

While current antidepressant use seemed to have some effect on weight reporting accuracy (Table 2), the associations were attenuated after adjusting for BMI. Household income, perceived stress, physical activity (total MET-h/week; MET = metabolic equivalent task), regular multivitamin use, gravidity, recency of last medical examination, smoking and alcohol were not associated with over- or under-reporting weight (data not shown).

Accuracy of self-reported height

Measured and self-reported height were highly correlated ($r = 0·96$); the average absolute difference between self-reported (CATI) and examiner-measured height was only 0·5 (sd 0·6; range 0–5·9) inches (1·3 (sd 1·5; range 0–15) cm). Slight variations between the CATI and examiner were likely due to different rounding conventions. Mean self-reported height was 64·6 (sd 2·6; range 50–75) inches (164·1 (sd 6·6; range 127·0–190·5) cm) and mean examiner-measured height was 64·7 (sd 2·5; range 50·7–75·1) inches (164·3 (sd 6·4; range 128·8–190·8) cm).

Over-reporting of height increased slightly with age and BMI. The odds of under-reporting height were higher among black women compared with whites. Also, women with less than a bachelor's degree had increased odds of misreporting their height compared with women with a bachelor's degree. No other factor was associated with differences in self-reported and measured height.

Accuracy of BMI based on telephone interview weight and height

The classification of overweight or obese BMI using self-reported measures was highly sensitive (0·95) and specific (0·96). For obese classification alone, sensitivity was 0·90 and specificity was 0·98.

BMI values based on CATI-reported and examiner-measured data were very close. The mean absolute difference between CATI-reported and examiner-measured BMI was only 0·7 (sd 0·8) kg/m²; the correlation was very high ($r = 0·98$; Fig. 4).

Among women with normal-range examiner-based BMI, BMI values calculated from CATI reports were within 4 % of measured BMI 83·4 % of the time (Table 4). However, despite an overall high correlation between BMI values from self-reported and examiner-measured data, there were noticeable discrepancies among women with lower and higher BMI. As shown, self-reported BMI was ≥5 % greater than measured BMI for about a quarter of underweight women. Also, BMI based on CATI-reported values was under-reported by ≥5 % for about 12 % of overweight women and 17 % of obese women.

Accuracy of self-completed questionnaire

Restricting to participants who completed both the questionnaire and CATI within 30 d of examiner assessment (n 13 985), the average absolute differences
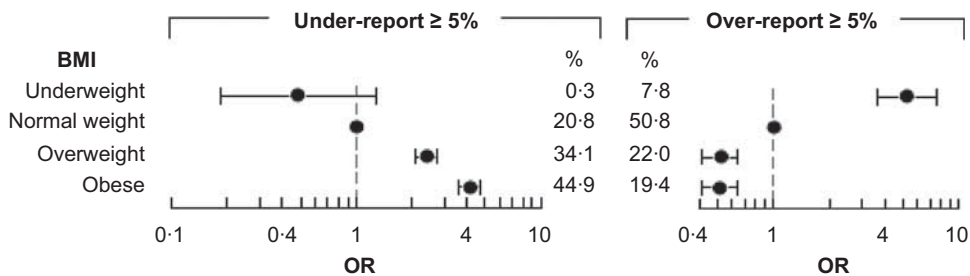
Fig. 2 The association between BMI and the accuracy of weight reported in a computer-assisted telephone interview: US women (*n* 18 639) aged 35–74 years, the Sister Study (2003–2009). Odds ratios were adjusted for age, race, education, perceived health status and marital status; 95 % confidence intervals are represented by error bars
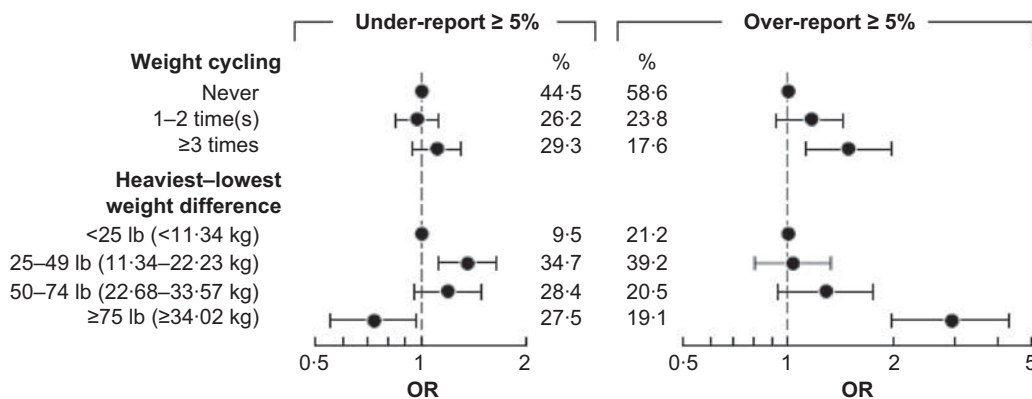

Fig. 3 The association between weight cycling and lifetime weight difference and the accuracy of weight reported in a computer-assisted telephone interview: US women (*n* 18 639) aged 35–74 years, the Sister Study (2003–2009). Odds ratios were adjusted for age, race, education, perceived health status, marital status and BMI; 95 % confidence intervals are represented by error bars
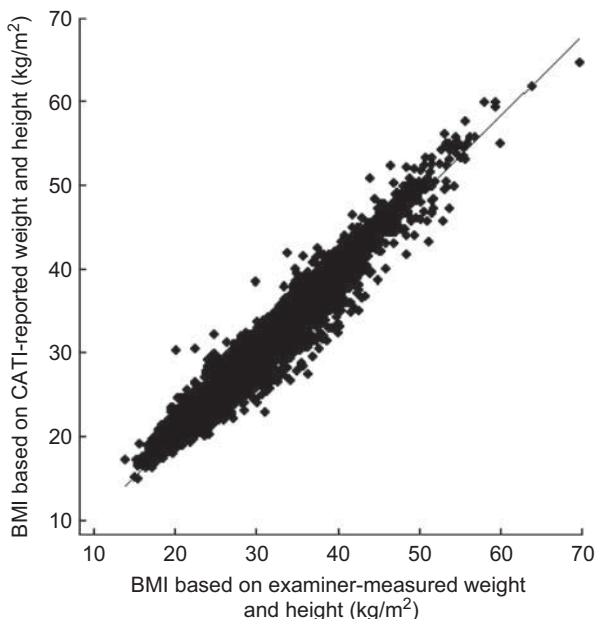
Fig. 4 Comparison of BMI calculated from weight and height reported in a computer-assisted telephone interview (CATI) *v.* examiner measures: US women (*n* 18 639) aged 35–74 years, the Sister Study (2003–2009); —— indicates fitted values

between CATI and measured height and weight were 0·4 (sd 0·6) inches (1·0 (sd 1·5) cm) and 3·2 (sd 4·0) lb (1·45 (sd 1·81) kg), respectively. The average absolute differences between questionnaire and measured height and weight were 0·5 (sd 0·6) inches (1·3 (sd 1·5) cm) and 3·4 (sd 3·6) lb (1·54 (sd 1·63) kg), respectively.

The tendency to under-report weight increased with BMI for both questionnaire and CATI although the differences were greater for telephone reports. For example, obese women were almost twice as likely to over-report by telephone compared with self-completed questionnaire (OR ratio = 1·86). Other differences were similarly magnified with telephone-reported data. Interestingly, while most trends suggested overweight women under-report their weight while underweight women over-report, women with large differences between heaviest and lowest weight also tended to over-report their weight when compared with examiner measurements, especially when reporting by telephone (see Appendix).

Reliability of self-reported weight and height

There were high correlations between the self-reported values for weight ($r = 0.99$) and height ($r = 0.98$).

Table 4 Percentage discrepancy between BMI based on CATI-reported values and examiner-measured values, by BMI (based on examiner-measured weight and height): US women (n 18 639) aged 35–74 years, the Sister Study (2003–2009)

BMI (based on examiner-measured weight and height)	n	Under-report BMI by $\geq5\%$ (row %)	Report BMI within 4 % (row %)	Over-report BMI by $\geq5\%$ (row %)
All women	18 634	10·5	83·4	6·2
Underweight ($<18\cdot5\,kg/m^2$)	241	0·8	72·6	26·6
Normal weight ($18\cdot5$–$24\cdot9\,kg/m^2$)	7600	5·5	86·6	7·9
Overweight ($25\cdot0$–$29\cdot9\,kg/m^2$)	5857	11·7	83·7	4·6
Obese ($\geq30\cdot0\,kg/m^2$)	4936	17·2	78·5	4·3

CATI, computer-assisted telephone interview.

The average absolute difference between weight reported in the CATI and questionnaire was 2·0 (SD 3·3; range 0–55) lb (0·91 kg (SD 1·50; range 0–24·95) kg). The absolute difference in height was 0·2 (SD 0·5; range 0–5) inches (0·5 (SD 1·3; range 0–12·7) cm). The absolute difference in time between self reports was 15 (SD 9) d. For weight, 80 % were within 3 lb (1·36 kg). For height, 99 % were within 1 inch (2·5 cm). The overall weighted κ was 0·95 for weight and 0·92 for height.

Factors associated with agreement in self-reported weight and height were largely similar to those for accuracy. Whereas height agreement decreased with age, weight agreement within 3 lb (1·36 kg) increased with age. Percentage agreement for weight and height increased with better perceived health status. Reporting agreement was inversely associated with BMI, weight cycling and lifetime weight difference. Findings were similar in analyses stratified by reporting order.

Discussion

Overall, women in the Sister Study reported weight and height accurately. Although participants were slightly leaner (on average 2 kg/m^2 lower in BMI) than middle-aged non-Hispanic white women in a smaller, nationally representative sample from the National Health and Nutrition Examination Survey (NHANES) 2003–2006[20], we confirmed previous findings that errors in reporting weight were associated with specific weight characteristics. Besides current weight status, we found that reporting accuracy was affected by excessive weight cycling (≥3 times) and extreme lifetime weight differences in adulthood (≥75 lb ($\geq34\cdot02$ kg)).

The present study is among the first to examine weight cycling and lifetime weight difference and reporting accuracy in a general population of women. Since weight cycling and lifetime weight difference both involve weight fluctuation, the extent to which the two variables were related was a concern. Weight cycling was associated with a lifetime weight difference of ≥30 lb ($\geq13\cdot61$ kg; χ^2 $P<0\cdot001$). However, 44 % of those who had a lifetime weight difference of ≥30 lb had never weight cycled, thus large changes in weight were not entirely explained by weight cycling.

Similar to previous studies, BMI values calculated from self-reported data were similar to those using measured data and there was high sensitivity for classifying a participant as overweight/obese or obese. Among adult women in the NHANES (1999–2004), there was substantial agreement between self-reported and measured BMI categories[7]. In an overweight Dutch sample, self-reported BMI was found to be reasonably accurate for the assessment of overweight/obesity prevalence[21]. Even with high correlation, there is still a potential for bias when examining associations between BMI based on self-reported measures and risk of disease and mortality[22]. Similar to our results, self-reported BMI in NHANES 2001–2006 and the National Health Interview Survey overestimated measured BMI values at the low end of the BMI scale (<22 kg/m^2) and underestimated values at the high end (>28 kg/m^2), and respondent sociodemographic characteristics were associated with some misclassification of obese people as overweight[13,23]. In our study, although BMI was underestimated by $\geq5\%$ for over 10 % of overweight and obese women, only 3 % of obese women under-reported their weight by $\geq10\%$ and fewer than 1 % of women in any BMI category under- or over-reported by $\geq15\%$. Furthermore, the average examiner-measured weight among obese women was 207 (SD 32) lb (93·90 (SD 14·52) kg) and the average amount under-reported by these women was only 3·3 (SD 6·8) lb (1·50 (SD 3·08) kg). Only 126 obese women under-reported by ≥20 lb ($\geq9\cdot07$ kg). For obese women, in particular, a 5 % difference in weight may have a negligible impact on associations with health outcomes.

Depression was of interest because it is associated with low self-esteem[24,25] and therefore could affect accuracy of weight reporting. However, diagnosis of depression or current use of antidepressant medication was not significantly associated with under- or over-reporting weight.

Several studies have suggested that respondents give more socially desirable answers in interviews than on self-administered questionnaires[26]. Despite finding a high correlation between CATI and questionnaire responses and seeing similar trends in accuracy for CATI and questionnaire, overweight and obese women reported weight more accurately on the questionnaire. While this finding might suggest that the anonymity of the self-completed

questionnaire promotes more honest reporting, it is also possible that women weighed themselves while completing the form at home. Access to a scale while completing the form may facilitate accurate reporting. Our participants may have been more motivated than others to do this because of the pending home visit during which they knew they would be weighed. Since women were asked to have their questionnaire ready for the examiner to collect, it is also possible that these forms were completed just before the home visit, increasing the likelihood of similar results. Thus our data may provide a 'best case' assessment of the validity of weight data reported on self-completed questionnaires.

Response rates and data quality can be higher in telephone interviews than mailed questionnaires[27,28]. CATI item non-response may have been minimized because interviewers asked each question, although women could refuse to answer. Having examiners physically collect the self-administered questionnaires may have helped reduce overall non-response for that form.

The current analysis has some unique caveats. Participants were told they would be weighed and measured during a home visit and therefore may have reported more accurately than they would have otherwise. Some variation between self-reported and measured weight may have occurred because examiners weighed women with clothing whereas women may have reported their weight without clothes. There was the potential for a learning effect caused by the order of the home visit and CATI self-report. Women who had the home visit first may have remembered their measured weight and height and later reported the same values in the CATI (59 % had home visit first; 37 % completed CATI first; 4 % completed both on same day). However, when we stratified the analyses by which measure came first, we found no evidence that the order of reporting influenced accuracy. Similarly, timing of the CATI in relation to filling out the questionnaire had little impact on reliability. Data were collected by many different examiners using different scales. Although all examiners were trained, we could not verify that measurement protocols were consistently followed.

Conclusions

US women in the Sister Study were reasonably reliable and accurate in reporting weight and height. Women with normal-range BMI reported most accurately. Overweight and obese women and those with fluctuations in their weight were less accurate, but even among obese women, few women under-reported their weight by ≥10 %. Nevertheless, even though self-reported and measured weight and height are highly correlated, bias can still exist in studies relying on self-reported data due to the tendency of overweight women to under-report and underweight women to over-report their weight. The present study is among the first to show that repeated weight cycling and large weight changes in adulthood are also associated with less accurate weight reporting in a general population of women.

Acknowledgements

References

1. Ogden CL, Carroll MD, Curtin LR *et al.* (2006) Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA* **295**, 1549–1555.
2. Engstrom JL, Paterson SA, Doherty A *et al.* (2003) Accuracy of self-reported height and weight in women: an integrative review of the literature. *J Midwifery Womens Health* **48**, 338–345.
3. Paradis AM, Perusse L, Godin G *et al.* (2008) Validity of a self-reported measure of familial history of obesity. *Nutr J* **7**, 27.
4. Gorber SC, Tremblay M, Moher D *et al.* (2007) A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev* **8**, 307–326.
5. Roberts RJ (1995) Can self-reported data accurately describe the prevalence of overweight? *Public Health* **109**, 275–284.
6. Bostrom G & Diderichsen F (1997) Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *Int J Epidemiol* **26**, 860–866.
7. Craig BM & Adams AK (2009) Accuracy of body mass index categories based on self-reported height and weight among women in the United States. *Matern Child Health J* **13**, 489–496.
8. Rowland ML (1990) Self-reported weight and height. *Am J Clin Nutr* **52**, 1125–1133.
9. Lahti-Koski M, Mannisto S, Pietinen P *et al.* (2005) Prevalence of weight cycling and its relation to health indicators in Finland. *Obes Res* **13**, 333–341.
10. Strohacker K & McFarlin BK (2010) Influence of obesity, physical inactivity, and weight cycling on chronic inflammation. *Front Biosci (Elite Ed)* **2**, 98–104.
11. Field AE, Byers T, Hunter DJ *et al.* (1999) Weight cycling, weight gain, and risk of hypertension in women. *Am J Epidemiol* **150**, 573–579.
12. White MA, Masheb RM, Burke-Martindale C *et al.* (2007) Accuracy of self-reported weight among bariatric surgery

candidates: the influence of race and weight cycling. *Obesity (Silver Spring)* **15**, 2761–2768.

13. Merrill RM & Richardson JS (2009) Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001–2006. *Prev Chronic Dis* **6**, A121.

14. Perez-Cueto FJ & Verbeke W (2009) Reliability and validity of self-reported weight and height in Belgium. *Nutr Hosp* **24**, 366–367.

15. Probst YC, Faraji S, Batterham M *et al.* (2008) Computerized dietary assessments compare well with interviewer administered diet histories for patients with type 2 diabetes mellitus in the primary healthcare setting. *Patient Educ Couns* **72**, 49–55.

16. McCabe RE, McFarlane T, Polivy J *et al.* (2001) Eating disorders, dieting, and the accuracy of self-reported weight. *Int J Eat Disord* **29**, 59–64.

17. Meyer C, Arcelus J & Wright S (2009) Accuracy of self-reported weight and height among women with eating disorders: a replication and extension study. *Eur Eat Disord Rev* **17**, 366–370.

18. Meyer C, McPartlan L, Sines J *et al.* (2009) Accuracy of self-reported weight and height: relationship with eating psychopathology among young women. *Int J Eat Disord* **42**, 379–381.

19. Centers for Disease Control and Prevention (2009) About BMI or Adults. http://www.cdc.gov/healthyweight/assessing/bmi/adult_BMI/index.html (accessed December 2009).

20. McDowell MA, Fryar CD, Ogden CL *et al.* (2008) *Anthropometric Reference Data for Children and Adults: United States, 2003–2006. National Health Statistics Reports* no. 10. Hyattsville, MD: National Center for Health Statistics.

21. Dekkers JC, van Wier MF, Hendriksen IJ *et al.* (2008) Accuracy of self-reported body weight, height and waist circumference in a Dutch overweight working population. *BMC Med Res Methodol* **8**, 69.

22. Keith SW, Fontaine KR, Pajewski NM *et al.* (2011) Use of self-reported height and weight biases the body mass index–mortality association. *Int J Obes (Lond)* **35**, 401–408.

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

24. Orth U, Robins RW & Meier LL (2009) Disentangling the effects of low self-esteem and stressful events on depression: findings from three longitudinal studies. *J Pers Soc Psychol* **97**, 307–321.

25. Orth U, Robins RW, Trzesniewski KH *et al.* (2009) Low self-esteem is a risk factor for depressive symptoms from young adulthood to old age. *J Abnorm Psychol* **118**, 472–478.

26. Okamoto K, Ohsuka K, Shiraishi T *et al.* (2002) Comparability of epidemiological information between self- and interviewer-administered questionnaires. *J Clin Epidemiol* **55**, 505–511.

27. Siemiatycki J (1979) A comparison of mail, telephone, and home interview strategies for household health surveys. *Am J Public Health* **69**, 238–245.

28. Brogger J, Bakke P, Eide GE *et al.* (2002) Comparison of telephone and postal survey modes on respiratory symptoms and risk factors. *Am J Epidemiol* **155**, 572–576.

Appendix

Comparing two types of self-reported weight (CATI and self-administered questionnaire) with examiner-measured weight among those who completed both self-reports within 30 d of the examiner measurement: US women aged 35–74 years (n 13 985), the Sister Study (2003–2009)

		Under-report by ≥5 % of examiner-measured weight				OR ratio	Over-report by ≥5 % of examiner-measured weight				OR ratio
	%	aOR$_{CATI}$	95 % CI	aOR$_{Questionnaire}$	95 % CI		aOR$_{CATI}$	95 % CI	aOR$_{Questionnaire}$	95 % CI	
BMI											
Underweight (<18·5 kg/m^2)	1·2	0·49	0·16, 1·56	0·55	0·17, 1·75	0·89	4·89	3·11, 7·69	3·06	2·22, 4·24	1·59
Normal weight (18·5–24·9 kg/m^2)	40·3	Ref.	–	Ref.	–	–	Ref.	–	Ref.	–	–
Overweight (25·0–29·9 kg/m^2)	31·7	2·44	2·05, 2·9	1·68	1·40, 2·01	1·45	0·48	0·37, 0·63	1·03	0·92, 1·15	0·47
Obese (≥30·0 kg/m^2)	26·8	3·97	3·34, 4·72	2·14	1·77, 2·58	1·86	0·52	0·40, 0·70	1·25	1·12, 1·40	0·42
Weight cycling											
Never	55·3	Ref.	–	Ref.	–	–	Ref.	–	Ref.	–	–
1–2 time(s)	25·1	0·97	0·82, 1·14	0·82	0·68, 1·00	1·17	1·11	0·85, 1·45	1·08	0·96, 1·20	1·03
≥3 times	19·6	1·10	0·91, 1·33	0·96	0·77, 1·2	1·15	1·32	0·95, 1·84	1·09	0·95, 1·25	1·21
Heaviest–lowest weight difference											
<25 lb (<11·34 kg)	16·7	Ref.	–	Ref.	–	–	Ref.	–	Ref.	–	–
25–49 lb (11·34–22·23 kg)	38·5	1·39	1·10, 1·75	1·29	1·01, 1·64	1·08	0·94	0·70, 1·26	1·06	0·92, 1·21	0·89
50–74 lb (22·68–33·57 kg)	24·5	1·25	0·96, 1·62	1·20	0·91, 1·59	1·04	1·21	0·83, 1·74	1·10	0·93, 1·29	1·10
≥75 lb (≥34·02 kg)	20·0	0·79	0·57, 1·11	0·92	0·64, 1·32	0·86	2·51	1·57, 4·02	1·30	1·06, 1·61	1·93

CATI, computer-assisted telephone interview; aOR, adjusted odds ratio (adjusted for age, race, education, perceived health status, marital status and BMI); Ref., referent category.