

Diagnostic in Obesity Comorbidities

A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review

S. Connor Gorber¹, M. Tremblay¹, D. Moher^{2,3} and B. Gorber⁴

¹Physical Health Measures Division, Statistics Canada, Ottawa, ON, Canada, ²Chalmers Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; ³Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; ⁴Independent Research Consultant, Ottawa, ON, Canada

Received 14 August 2006; revised 25 October 2006; accepted 30 October 2006

Address for correspondence: S Connor Gorber, Physical Health Measures Division, Research and Analysis, Statistics Canada, Main Building Room 0005, 150 Tunney's Pasture Driveway, Ottawa, ON, K1A 0T6, Canada.
E-mail: sarah.connorgorber@statcan.ca

Summary

Obesity is a rapidly increasing public health problem, with surveillance most often based on self-reported values of height and weight. We conducted a systematic review to determine what empirical evidence exists regarding the agreement between objective (measured) and subjective (reported) measures in assessing height, weight and body mass index (BMI). Five electronic databases were searched to identify observational and experimental studies on adult populations over the age of 18. Searching identified 64 citations that met the eligibility criteria and examined the relationship between self-reported and directly measured height or weight. Overall, the data show trends of under-reporting for weight and BMI and over-reporting for height, although the degree of the trend varies for men and women and the characteristics of the population being examined. Standard deviations were large indicating that there is a great deal of individual variability in reporting of results. Combining the results quantitatively was not possible because of the poor reporting of outcomes of interest. Accurate estimation of these variables is important as data from population studies such as those included in this review are often used to generate regional and national estimates of overweight and obesity and are in turn used by decision makers to allocate resources and set priorities in health.

Keywords: Height, measurement error, self-report, weight.

obesity reviews (2007) **8**, 307–326

Introduction

Height and weight are important indicators of population health because of their role in calculating the body mass index (BMI), a common measure of obesity (1). Obesity rates have been on the rise in many countries over the last several decades, and more than doubled in Canada over the 13-year period from 1985 to 1998 (2). Internationally, the World Health Organization has declared that obesity has reached epidemic proportions (3) and that it will be one of the greatest challenges and risk factors for chronic disease in this century (4). Recent evidence indicates that the prevalence of obesity in Canada is 23% in adults (5)

and 8% in children (6). The health costs of obesity in Canada represent 2.2% of total healthcare costs (7).

Self-report measures are one of the most common ways to collect data on height and weight and have consistently been shown to be associated with mortality and morbidity (8–10). Self-reports have the advantages of practicality and low cost, they are quick and easy to administer and are a good method for sampling large numbers of individuals. However, the tools are subject to certain limitations: questions may be misunderstood, participants may not accurately recall past events and response bias e.g. social desirability or response acquiescence are not uncommon (11).

Social desirability is particularly problematic with sensitive questions (e.g. about weight or obesity) and can be influenced by the method of data collection (12–14). For example, in the Canadian Community Health Survey, one of Canada's largest national health surveys (15), the proportion of respondents reporting being obese (based on self-reported values of height and weight) was estimated to be 18% for those who were interviewed in person compared with an estimate of only 13% for those who were asked the same questions by phone (16). These inconsistencies can be even more pronounced when for operational and budgetary reasons survey modes do not stay consistent over time. Similarly, changes in question wording or respondent characteristics among surveys, or within a survey over time, also have the potential to impact the quality of self-reported estimates. These issues make it difficult to get accurate self-reported estimates of height, weight and obesity over time. Compounding the problem is the fact that the influence of social desirability on self-reports has the potential to change over time as social and cultural norms about weight and obesity change.

Because of the limitations associated with self-report measures, objective or direct measures have been recommended to improve measurement precision. Despite the benefits, the costs of this type of measurement are often prohibitively high, particularly in population research where the samples are large. The intrusive nature of the measures also has the potential to impact response rates and contribute to attrition or bias. As well, such measures often require special training of interviewers for accurate administration and classification.

In order to ensure that researchers are employing the most appropriate tools in their research and to ensure that decision making is based upon the best available evidence, the degree of accuracy of self-reported measures needs to be determined. We conducted this systematic review to determine what empirical evidence exists regarding the concordance of objective and subjective measures in assessing height and weight. The primary objective was to compare direct vs. self-report measures for assessing height and weight in observational and experimental studies of adult populations. A secondary objective was to examine how the use of self-report and direct measures of height and weight impact on estimates of the prevalence of obesity and overweight, as measured by the BMI. Although it is recognized that BMI is only one of many methods used to measure obesity, it is employed in this review because it can be directly calculated from weight and height measurements.

Criteria for considering studies for this review

Any observational or experimental studies that included a direct comparison of self-report and objective measures for

the outcomes of interest were eligible for inclusion. All study designs were eligible (e.g. retrospective, prospective, case control) and both published and unpublished literature was examined.

Only studies on adult populations over the age of 18 years were considered. Children were excluded as they are still in an important stage of development where variables such as weight may change over short periods of time. It has also been suggested that reporting error in children and adolescents may be of a different nature than that in adults (17).

Any form of weight or height measurement whether by portable or balance beam scale, stadiometer, tape measure or other technique was eligible for inclusion. Likewise, any form of self-reporting technique could be included (e.g. interviewer administered, self-complete). Proxy reporting (when one individual responds for another) was not eligible for inclusion.

Search strategy for identification of studies

The following electronic bibliographical databases were searched:

- MEDLINE – from 1966 to January Week 4 2006;
- EMBASE – from 1980 to Week 4 2006;
- CINAHL – from 1982 to December Week 2 2005;
- PsycINFO – from 1806 to January Week 4 2006;
- Health and Psychosocial Instruments (HAPI) – from 1985 to September 2005;
- SPORTDiscus – from 1830 to January 2006.

The search strategy used for searching MEDLINE is provided as an example (see Table 1) but was modified according to the indexing systems of other databases. No language of publication restrictions was imposed. The Ovid interface was used for all electronic searches. Grey litera-

Table 1 MEDLINE search strategy

1	objective measure\$.mp.
2	direct measure\$.mp.
3	physical measure\$.mp.
4	measured weight.mp.
5	measured height.mp.
6	exp Self Concept/
7	self report\$.mp.
8	body weight/or body weight changes/or overweight/or thinness/
9	Body Height/
10	height.mp.
11	weight.mp. or 'Weights and Measures'/
12	1 or 2 or 3 or 4 or 5
13	6 or 7
14	8 or 9 or 10 or 11
15	12 and 13 and 14
16	limit 15 to humans

ture, such as published abstracts/proceedings, published lists of theses and dissertations, and government reports were also identified where possible.

Methods of the review

Selecting citations

Titles and abstracts of potentially relevant citations were first reviewed to assess adherence to the inclusion criteria. The full text of all articles that met the criteria was then further screened by two independent reviewers (SCG, BG). Searches of the bibliographies of texts were also performed in order to identify additional studies, which were subsequently retrieved. The two reviewers examined the full text versions of remaining citations and selected those that met the inclusion criteria for the study. There were no disagreements about inclusion.

Data extraction

Standardized data extraction forms were completed by one reviewer (SCG) and verified by the other (BG). Information was extracted on the type of study, participant characteristics, sample size, methods of weight, height and BMI measurement, time between self-report and direct measurement, measurement setting, tools and equipment, order of measures, and results (means and mean differences between measures, measures of variance). Reviewers were not blinded to the authors or journals when extracting data. The data extraction form is found in Appendix I.

Quality assessment

Because both experimental and observational study designs were included in this review, two tools were used to assess quality. Jadad's scale (18) was used to assess the quality of randomized controlled trials and the Downs and Black (19) tool was used to assess the quality of non-randomized designs. The Jadad scale includes three items assessing the quality of randomization generation, blinding and drop-outs and withdrawals. The scale ranges from 0 to 5 with higher scores indicating higher quality. The Downs and Black instrument was one of six instruments recommended (20) for use in systematic reviews of non-randomized studies. The Downs and Black instrument was selected as it contained the highest number of relevant items for the needs of this review. However, as not all items were relevant to the studies included in this review (many of which were based on surveys), a modified version of the checklist was employed with the following items omitted: items 5 and 8 in the reporting scale, items 11, 15 and 19 in the section on bias, items 21–26 relating to confounding and item 27 addressing power. The final checklist was made up of 15

items with a maximum score of 15 points (with higher points indicating superior quality) rather than the original 32 points. The checklist covered the categories of reporting, external validity and internal validity (bias).

Results

Description of studies

A total of 328 citations were identified based on the preliminary search of electronic databases, reference lists and grey literature (see Fig. 1). Of these 137 were identified in MEDLINE, an additional 35 in EMBASE, 117 in CINAHL, 16 in PsycINFO and one in SPORTDiscus. Searching HAPI did not uncover any citations that had not been previously identified. Searching bibliographies of key

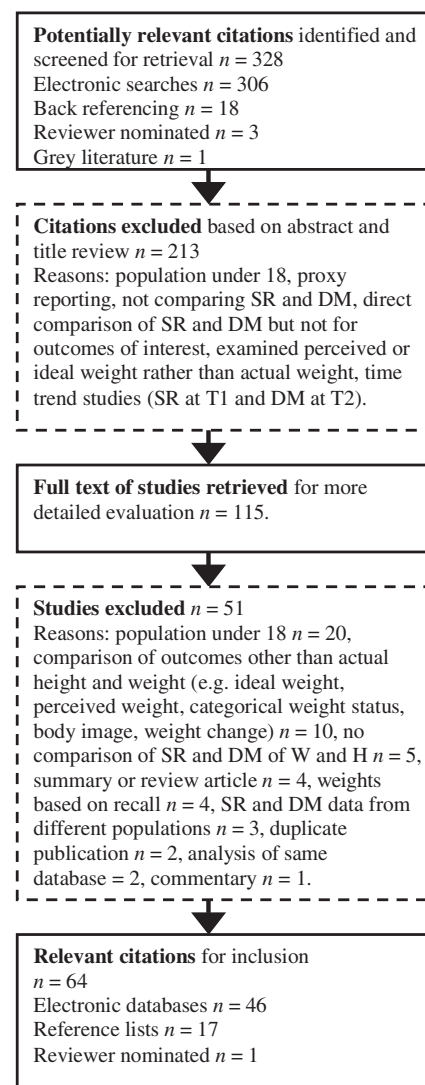


Figure 1 Results of the literature search. DM, direct measure; H, height; SR, self-report; W, weight.

articles, mostly Engstrom (21) who had conducted a review examining the concordance between self-reported and directly measured height and weight in women yielded an additional 18 citations. The three reviewer nominated citations were discovered during searching. The one citation that was included from the grey literature search was published by Statistics Canada.

In total, 115 full text articles were retrieved for detailed assessment. Of these 64 met the criteria for study inclusion (see Table 2). Common reasons for excluding studies included a focus on outcomes other than height and weight, a focus on children or adolescents,¹ not having both directly measured and self-reported data within the same study on the same population and examining ideal or perceived weight or height rather than actual values.

Four of the articles (17,22–24) analysed data from the same database, the National Health And Nutrition Examination Survey (NHANES), three of which looked at the data from the NHANES III cohort collected between 1988 and 1994. NHANES is 'a program of studies designed to assess the health and nutritional status of adults and children in the United States' (25) (p. 2). In order to prevent double counting, only one of the NHANES III studies (22) was retained. Kuczmarski was selected because it had the most comprehensive outcome reporting (it provided mean differences with corresponding measures of variance). Rowland's (17) data were based on the NHANES II cohort collected from 1976 to 1980, and therefore, there are no concerns about overlap.

Also requiring clarification is the inclusion of Stunkard and Albaum (26) and Schlichting *et al.* (27), both published in 1981. Stunkard examined data that were collected on subjects at eight different medical and non-medical sites in two countries (Denmark and the United States). Schlichting's study examined the same dataset but focused on analysis of the Danish data only. Hence, Stunkard's mean difference values for weight were used in this review, because they were more comprehensive, but because they did not report any results on height, Schlichting's data for the height outcome were also included.

Of the 64 studies that were retained, 60 used different types of observational designs (17,22,26–83) (e.g. case control, cross-sectional), two used experimental designs (84,85), one was a letter (86) and there was one unpublished dissertation (87). Two studies were in languages other than English (Portuguese (81) and Spanish (82)). Studies that were included were published over a 26-year period between 1979 and 2005. The primary aim of the articles differed but each compared direct measures of height or weight with their self-reported counterparts.

¹The database search was not restricted to adults, as this would have excluded some studies (e.g. LeJarraga *et al.* (28)) that examined the characteristics of both children and their parents.

Participants in the studies ranged from 10 to over 60 years, with the oldest participant reported to be 89 years old. Although the focus of the review is on those aged 18 years and over, studies that had a range of ages less than 18 years were not excluded as long as the mean age of the majority of the population was over 18 years. The study by Allison *et al.* (29), for example, had an age range of 10–67 years but a mean age of 38.5 ± 12.0 years. A decision was made to include this study because only a small number of cases fell below the 18-year cut-off. Sample size ranged from a low of 44 in the dissertation (87) to a high of over 16 000 in Kuczmarski's NHANES analysis (22).

The majority of studies reported which tools were used to measure the weight and height of participants in the various studies, but 23 studies provided no information about what equipment was used to take the measurements (the two non-English language publications were not assessed on these characteristics). When data on measurement tools were reported, height was most commonly measured by stadiometer, anthropometer, or some type of measuring tape or ruler, and a variety of scales (balance beam, digital, or portable) were used to measure weight. Most studies provided information about what participants wore during the measurements, which was usually light indoor clothes and no shoes. In six studies (30,47,52,59,60,61), participants wore only undergarments and in another six, they were measured and weighed in hospital gowns (17,22,55,57,66,75). In one study focused on height (44), participants were purposely given no instructions about whether shoes should be worn for the measurement as authors wanted to capture how the participants presented themselves to the world. Sixteen studies did not describe the clothing worn by participants.

Self-report data were most often collected by questionnaire (either by mail, telephone or most often in person) with 16 studies collecting data in an interview. Self-reports were conducted prior to the direct measurement in over 80% of studies. In two studies (36,43), the sample was divided into two groups with one undergoing the direct measurement before the self-report and the other completing the self-report first. Only one study (49) consistently measured participants before the self-report was completed. Five studies (42,53,61,68,83) did not provide enough information to determine which came first. Twenty-five studies collected the direct measurement with the captured self-report immediately afterwards, with Roberts (66) reporting the longest delay between measures (3–8 months).

Quality assessment

Results of the quality assessment are presented in Table 3. Quality was assessed on 60 of the studies. Zhang's letter (86) was not assessed, because although it presented data, it did not provide enough information to

Table 2 Study and participant characteristics

ID	Study – first author (reference)	Sample size	Age range or (mean ± SD)	Population	Location	Outcomes	Time lag	Measurement order
1	Allison 1998 (29)	104	10–67	Obese	USA	H, W, BMI	None	SR first
2	Alvarez-Torices 1993 (30)	572	18+	General population	Spain	H, W, BMI	NR	SR first
3	Avila-Funes 2004 (31)	1 707	24–95	Adult population	Mexico	H, W, BMI	None	SR first
4	Black 1998 (85) ^F	223	18–82	General population	USA	W	None	SR first
5	Bolton-Smith 2000 (32)	1 836	25–64	Patients from GP registers	Scotland	H, W, BMI	2 weeks	SR first
6	Booth 2000 (33)	1 140	18–78	Adults	South Australia	H, W, BMI	2 weeks	SR first
7	Bostrom 1997 (34)	3 208	18–84	Adults	Sweden	H, W, BMI	4–6 months	SR first
8	Brown 2002 (35)	409	19–67	Adults	USA	H	None	SR first
9	Brunet 2003 (87) ^{DU}	44	18–23	Female university athletes	USA	H, W	None	SR first
10	Cash 1989 (36)	133	18–52	Female college students	USA	W	10 min	Group 1: SR first; group 2: DM first
11	Chor 1999 (37)	322	38.5	Bank employees	Brazil	H, W, BMI	None	SR first
12	Clemente 2004 (81) ^T	380	18–30	University students	Portugal	H, W, BMI	NA	NA
13	de Araujo 2003 (58)	193	16–73*	Attendees at cardiology congress	Brazil	H, W, BMI	None	SR first
14	DelPrete 1992 (38)	82	18+	Adult – weight loss programme participants	USA	H, W	2.5 weeks	SR first
15	Doll 1998 (39)	102 BN 204 HC	16–35	General population, BN and healthy	UK	H, W, BMI	NR	SR first
16	Fonseca 2004 (40)	3 713	22–70	University employees	Brazil	H, W, BMI	None	SR first
17	Forster 1988 (41)	28	NR	University employees from worksite weight control programme	USA	W	NR	SR first
18	Giles 1991 (42)	7 999	Male: 22.2 ± 4.6 Female: 23.1 ± 5.4	US army personnel	USA	H	NR	NR
19	Gunnell 2000 (43)	257	56–78	Elderly	England, Scotland	H, W, BMI	NR	Half prior and half after
20	Hensley 1998 (44)	59	20.3	College students	USA	H	None	SR first
21	Hill 1998 (45)	2 258	16–64	Family Health Services Authority register	England	H, W, BMI	1–4 months	SR first
22	Imrhan 1996 (46)	469	18–46	College students	USA	H, W	None	SR first
23	Jacobson 2001 (47)	62	Mean 19.9	College students	USA	H, W, BMI	None	SR first
24	Jalkanen 1987 (48)	1 1880	30–64	General population	Finland	W	10 days	SR first
25	Kinney 1988 (49)	167	Mean 59.0 ± 13.6	Male Veterans Administration patients	USA	W	1 month	DM first

Table 2 Continued

ID	Study – first author (reference)	Sample size	Age range or (mean ± SD)	Population	Location	Outcomes	Time lag	Measurement order
26	Klag 1993 (50)	78	49–75	Physicians	USA	H, W, BMI	None	SR first
27	Kuczmarski 2001 (22)	16 573	20+	General population	USA	H, W, BMI	2–4 weeks	SR first
28	Kuskowska-Wolk 1989 (52)	301	16–84	Patients presenting for medical appointment	Sweden	H, W, BMI	None	SR first
29	Kuskowska-Wolk 1992 (51)	3 390	18–84	Adults	Sweden	H, W, BMI	4–6 months	SR first
30	Lackland 1990 (53)	3 202	NR	Adults	USA	H, W, BMI	NR	NR
31	Larson 2000 (54)	56	18–59	Healthy	USA	H, W	1 week	SR first
32	Lawlor 2002 (55)	1 310	60–79	Women	England, Scotland, Wales	W	6 weeks	SR first
33	LeJarraga 1995 (28)	144	20–66	Parents of children attending growth clinics	Argentina	H	None	SR first
34	Masheb 2001 (56)	108	21–61	Adults with binge eating disorder	USA	W	None	SR first
35	McCabe 2001 (57)	81 ED, 163 D and ND controls	Mean range 20.0–26.1	Hospital eating disorders programme (ED), college students (controls)	Canada	H, W	2–5 days	SR first
36	Nakamura 1999 (59)	354	20–42	Female employees of a computer assembly factory	Japan	H, W, BMI	1 week	SR first
37	Nawaz 2001 (60)	95	30–65	Overweight or obese women	USA	H, W, BMI	2 weeks	SR first
38	Niedhammer 2000 (61)	7 350	35–50	Employees of Electricité de France-Gaz de France	France	H, W, BMI	<6 months	NR
39	Nieto-Garcia 1990 (62)	7 455	20–79	60% hyperlipidemic	North America	H, W, BMI	None	SR first
40	O'Connell 2005 (84) ^f	201	18–40	Women with regular menstrual cycles	USA	W	None	SR first
41	Palta 1982 (63)	1 344	30–69	Diastolic blood pressure >95 mm HG	USA	H, W	3 weeks	SR first
42	Payette 2000 (64)	475	Mean 81.8	Cognitively intact and impaired elderly	Canada	H, W	None	SR first
43	Pirie 1981 (65)	3 407	20–59	Caucasians	USA	H, W	1 month	SR first
44	Quiles 1996 (82) [†]	1 387	15+	General population	Spain	W, H, BMI	NA	NA
45	Roberts 1995 (66)	1 622	18–64	Adults	Wales	H, W	3–8 months	SR first
46	Rona 1989 (67)	68 couples	NR	Afro-Caribbean, Asian and Caucasian couples	Britain	H	NR	SR first
47	Rowland 1990 (17)	11 284	20–74	Adults	USA	H, W	2–6 weeks	SR first
48	Santillan 2003 (68)	961	Mean 42 ± 14	Asthmatic outpatients, workers and families	Mexico	H, W, BMI	NR	NR

Table 2 Continued

ID	Study - first author (reference)	Sample size	Age range or (mean ± SD)	Population	Location	Outcomes	Time lag	Measurement order
49	Schlichting 1981 (27)	752	16-66	Insurance applicants	Denmark	H, W	NR	SR first
50	Schmidt 1993 (69)	659	15-64	Adults	Brazil	W	NR	SR first
51	Smith 1992 (70)	103	Mean 23.1 ± 6.7	Undergraduate women, two with BN	USA	W	None	SR first
52	Spencer 2002 (71)	4 808	35-76	Middle-aged adults	Britain	H, W, BMI	few weeks	SR first
53	Stewart 1982 (72)	3 373	14-61	Families	USA	H, W	<3 months	SR first
54	Stewart 1987 (73)	1 523	35-65	Adult Caucasians	New Zealand	H, W	None	SR first
55	Stunkard 1981 (26)	1 302	Mean 40.0 ± 0.6 (USA), mean range 30.2-47.5 (Denmark)	Insurance applicants (Denmark), broad range of persons (USA)	Denmark, USA	W	NR	SR first
56	Tell 1987 (74)	146	28-63	Former weight control programme participants	USA	W	2-64 days	SR first
57	Tienboon 1992 (75)	213 families 135 F, 190 M	Mean 44.8 ± 5.8 F Mean 41.8 ± 5.3 M	Parents of 15-year-olds	Australia	H, W, BMI	None	SR first
58	Vailas 1998 (76)	131	60 and over	Meal programme participants	USA	H, W	None	SR first
59	van der Voort 2000 (77)	1 155	50-80	Postmenopausal women	Netherlands	H	None	SR first
60	Wada 2005 (78)	5 401	35-64	Public servants	Japan	H, W, BMI	NR	SR first
61	Wang 1997 (83)	2 731	NR	Females	England	H	NR	NR
62	Wing 1979 (79)	S1:78 S2:11	NR 18-89	Psychology undergraduate students	USA	H, W	None	SR first
63	Zhang 1993 (86) [†]	8	40-54	Attendees at health fair	USA	H, W	NR	SR first
64	Ziebland 1996 (80)	2 205	35-64	Caucasian perimenopausal women	Britain	H, W	NR	SR first

^{*}This was the range for a larger sample of 844; this study includes only a subsample of these cases.

[†]Portuguese.

[‡]Spanish.

[§]dissertation; [¶]experimental design; [‡]letter; [¶]unpublished.

BMI, body mass index; BN, bulimia nervosa; D, dieter; DM, direct measure; ED, eating disorder (anorexia or bulimia nervosa); F, fathers; GP, general practitioner; H, height; HC, healthy controls; M, mothers; NA, not assessed; ND, non-dieter; NR, not reported; S, study; SR, self-report; W, weight.

Table 3 Quality assessment of studies

Observational studies: assessed with Downs and Black quality assessment tool (19)					
ID	Study – first author (reference)	Quality scores (value)			
		Reporting (8)	External validity (3)	Internal validity (4)	Total (15)
1	Allison 1998 (29)	7	1	4	12
2	Alvarez-Torices 1993 (30)	8	3	4	15
3	Avila-Funes 2004 (31)	8	2	4	14
5	Bolton-Smith 2000 (32)	7	2	3	12
6	Booth 2000 (33)	5	2	4	11
7	Bostrom 1997 (34)	6	2	4	12
8	Brown 2002 (35)	7	1	4	12
9	Brunet 2003 (87)	6	1	4	11
10	Cash 1989 (36)	5	0	3	8
11	Chor 1999 (37)	6*	1	4	11
13	de Araujo 2003 (58)	6	1	4	11
14	DelPrete 1992 (38)	7	3	4	14
15	Doll 1998 (39)	7	1	3	11
16	Fonseca 2004 (40)	7	2	4	13
17	Forster 1988 (41)	6	2	3	11
18	Giles 1991 (42)	5	0	3	8
19	Gunnell 2000 (43)	8	2	3	13
20	Hensley 1998 (44)	5	1	4	10
21	Hill 1998 (45)	4	2	3	9
22	Imrhan 1996 (46)	5	1	4	10
23	Jacobson 2001 (47)	7	1	4	12
24	Jalkanen 1987 (48)	7	2	4	13
25	Kinney 1988 (49)	7	1	4	12
26	Klag 1993 (50)	8	1	4	13
27	Kuczmariski 2001 (22)	6	2	3	11
28	Kuskowska-Wolk 1989 (52)	8	3	4	15
29	Kuskowska-Wolk 1992 (51)	7	2	3	12
30	Lackland 1990 (53)	3	0	3	6
31	Larson 2000 (54)	6	0	3	9
32	Lawlor 2002 (55)	8	2	4	14
33	LeJarraga 1995 (28)	7	2	4	13
34	Masheb 2001 (56)	7	1	3	11
35	McCabe 2001 (57)	7	1	4	12
36	Nakamura 1999 (59)	7	1	4	12
37	Nawaz 2001 (60)	8	0	4	12
38	Niedhammer 2000 (61)	7	2	3	12
39	Nieto-Garcia 1990 (62)	5*	2	3	10
41	Palta 1982 (63)	6	1	4	11
42	Payette 2000 (64)	8	1	4	13
43	Pirie 1981 (65)	6	2	4	12
45	Roberts 1995 (66)	5*	0	3	8
46	Rona 1989 (67)	7	1	3	11
47	Rowland 1990 (17)	7	3	3	13
48	Santillan 2003 (68)	5	0	3	8
49	Schlichting 1981 (27)	5	1	3	9
50	Schmidt 1993 (69)	7	1	4	12
51	Smith 1992 (70)	8	1	4	13
52	Spencer 2002 (71)	7	1	4	12
53	Stewart 1982 (72)	6	2	3	11
54	Stewart 1987 (73)	7	3	4	14
55	Stunkard 1981 (26)	5	1	3	9
56	Tell 1987 (74)	7	1	4	12
57	Tienboon 1992 (75)	7	1	4	12
58	Vailas 1998 (76)	7	2	4	13
59	van der Voort 2000 (77)	7*	1	4	12
60	Wada 2005 (78)	6	3	3	12
62	Wing 1979 (79)	5	1	3	9
64	Ziebland 1996 (80)	5*	0	3	8

Table 3 Continued

Experimental studies: assessed with Jadad quality assessment tool (18)		
		Quality score (max 5)
4	Black 1998 (85)	1
40	O'Connell 2005 (84)	3
Not assessed		
		Reason
12	Clemente 2004 (81)	foreign language
44	Quiles 1996 (82)	foreign language
61	Wang 1997 (83)	not enough data provided to assess quality
63	Zhang 1993 (86)	not enough data provided to assess quality

*These studies reported no *P*-values, and were therefore given a score of 0 for item 10 on reporting probability values.

determine a quality assessment. Similarly, Wang (83), which was mostly a review article with some new analyses presented, lacked information for an appropriate quality score to be calculated. Quality scores were also not computed for the two non-English language articles. Of the articles that remained, 58 were assessed with the Downs and Black checklist and two using the Jadad scale. The range on the Downs and Black tool was 6 to 15 (maximum possible score was 15) with a mean of 11.4 ± 1.9 .

Citations were assessed based on the degree to which they presented data relating to the outcomes studied in this review (e.g. height and weight) even if these were only secondary outcomes in the studies. Common gaps in reporting included not clearly describing the intervention of interest (e.g. the method of height or weight measurement), not presenting simple outcome data such as the mean weight or height in the self-reported and measured groups so that major analyses and conclusions could be verified and not reporting the associated measures of random variability. Although most studies carried out some sort of significance testing, many did not report the actual probability values, which resulted in the deduction of a point.

Thirteen per cent of studies received no points in the external validity category while 57% received one point or less. Most of the point losses were due to a lack of representativeness of the samples, with participants who were asked or those who agreed to participate not being representative of the entire population from which they were recruited. No study lost more than one point in the internal validity checklist, with over half receiving perfect scores in this category. Most deductions resulted from not adjusting for different lengths of follow-up of subjects in the analyses (e.g. if the delay between self-report and direct measures was excessively long – over 4 weeks – and was not consis-

tent from subject to subject). Points were also deducted on item 18 when mean differences were not calculated or could not be calculated from the data.

The two experimental studies were assessed with the Jadad scale. Black *et al.* (85) received one point for describing the study as randomized and O'Connell *et al.* (84) received one point for describing the study as randomized, one point for describing withdrawals and dropouts and an additional point because the method of randomization (random numbers table) was an appropriate randomization technique.

Assessing quality, particularly in observational study designs, is not without flaws. Even the items on the Downs and Black checklist, which was developed for non-randomized designs, did not always apply to the studies in this review, and at times the interpretation of items had to be modified in order to apply them to the included studies.

Height

Fifty-three studies contained data on self-reported and measured height and these are displayed in Table 4. The table presents the mean differences between self-reported and measured height (SR-DM) for the overall sample as well as the breakdowns for males and females (in some cases, the actual mean difference was provided and in others it was calculated by subtracting the directly measured value from the self-reported value). What is most apparent from examining the table is the lack of data that are presented on the key outcomes, even in many studies where the goal was to compare measured and reported height. Only 28 of the studies provided both mean differences (either by sex or overall) and a corresponding measure of variance, which are required for pooling data.

Table 4 Mean differences in height (self-report–direct measure)

ID	Study – first author (reference)	Mean difference (cm)					
		Total	SD (95% CI)	Men	SD (95% CI)	Women	SD (95% CI)
General population							
2	Alvarez-Torices 1993 (30)	2.2	(1.9, 2.6)	2.3	(1.9, 2.7)	2.2	(1.6, 2.7)
3	Avila-Funes 2004 (31)	1.7	–	1.2	–	2.2	–
5	Bolton-Smith 2000 (32)	–	–	–1.3	2.5	–1.7	2.4
6	Booth 2000 (33)	–	–	–	–	–	–
7	Bostrom 1997 (34)	0.7	(0.6, 0.8)	0.6	–	0.8	–
8	Brown 2002 (35)	0.6	–	–	–	–	–
27	Kuczmarowski 2001 (22) 20–59 years	–	–	1.0	0.1	0	0.1
29	Kuskowska-Wolk 1992 (51)	–	–	0.5	–	1.0	–
30	Lackland 1990 (53)	–	–	–	–	–	–
31	Larson 2000 (54)	0.9	–	–	–	–	–
44	Quiles 1996 (82)	1.0	–	0.6	–	1.4	–
45	Roberts 1995 (66)	–	–	1.4	2.3	0.7	2.8
47	Rowland 1990 (17)	–	–	1.4	2.6	0.6	2.8
52	Spencer 2002 (71)	–	–	1.2	2.6	0.6	2.7
54	Stewart 1987 (73)	1.9	(1.8, 2.1)*	2.1	–	1.6	–
61	Wang 1997 (83)	–	–	–	–	–	–
64	Ziebland 1996 (80)	–	–	–	–	–	–
	BMI < 20	–	–	0.6	–	0.4	–
	BMI 20–24	–	–	0.9	(0.6, 1.2)	0.9	(0.5, 1.2)
	BMI 25–29	–	–	1.0	(0.7, 1.2)	1.1	(0.9, 1.4)
	BMI > 30	–	–	1.6	(1.0, 2.1)	1.6	(1.2, 2.0)
Overweight or weight loss participants							
1	Allison 1998 (29)	1.5	–	1.7	–	1.5	–
14	DelPrete 1992 (38)	1.8	2.7	3.0	2.5	1.3	2.6
37	Nawaz 2001 (60)	–	–	–	–	–	–
	25 ≤ BMI < 30	NA	NA	NA	NA	0.1	0.4
	30 ≤ BMI ≤ 35	NA	NA	NA	NA	0	0.9
	35 < BMI ≤ 40	NA	NA	NA	NA	0.3	1.0
	BMI > 40	NA	NA	NA	NA	0.9	0.9
Special populations							
13	de Araujo 2003 (58)	–	–	1.0	–	0.6	–
21	Hill 1998 (45)	7.5	(4.8, 10.0)	–	–	–	–
26	Klag 1993 (50)	1.4	0.3†	–	–	–	–
28	Kuskowska-Wolk 1989 (52)	–	–	0.9	–	1.9	–
39	Nieto-Garcia 1990 (62)†	0.6	(0.6, 0.6)	0.9	(0.9, 1.0)	0.3	(0.3, 0.4)
41	Palta 1982 (63)	–	–	2.3	2.7	0.9	2.4
42	Payette 2000 (64)	5.0	–	5.0	–	5.0	–
43	Pirie 1981 (65)	–	–	0.2	–	–0.3	–
48	Santillan 2003 (68)	–	–	–	–	–	–
49	Schlichting 1981 (27)	–	–	–	–	–	–
58	Vailas 1998 (76)	2.0	3.6	1.5	3.8	2.3	3.6
Parents/families							
33	LeJarraga 1995 (28)	–	–	–	–	–	–
	Private growth clinic	–	–	0.9	2.0	1.1	1.9
	Public growth clinic	–	–	–0.3	2.6	0.9	3.1
46	Rona 1989 (67)	–	–	–	–	–	–
	Afro-Caribbean	–	–	0.2	6.5	2.9	11.2
	Asian	–	–	0.6	4.2	4.5	6.2
	Caucasian	–	–	3.9	8.8	–0.1	2.7
53	Stewart 1982 (72)	1.7	1.0	–	–	–	–
57	Tienboon 1992 (75)	–	–	2.9	2.2	1.7	2.3
Eating disorders							
15	Doll 1998 (39)	–	–	–	–	–	–
	Healthy controls	0.2	(0.1, 0.5)	–	–	–	–
	Bulimia nervosa	0	(–0.5, 0.4)	–	–	–	–

Table 4 Continued

ID	Study – first author (reference)	Mean difference (cm)					
		Total	SD (95% CI)	Men	SD (95% CI)	Women	SD (95% CI)
35	McCabe 2001 (57)						
	Anorexics	0.9	2.0	–	–	–	–
	Bulimics	1.1	2.0	–	–	–	–
	Dieters	1.7	2.3	–	–	–	–
	Non-dieters	2.0	2.6	–	–	–	–
Students							
9	Brunet 2003 (87)	NA	NA	NA	NA	–	–
12	Clemente 2004 (81)	–	–	1.2	2.2	2.7	2.3
20	Hensley 1998 (44)	–	–	–0.7	–	–0.9	–
22	Imrhan 1996 (46)	–	–	–	–	–	–
23	Jacobson 2001 (47)	–	–	1.0	–	0	–
62	Wing 1979 (79)	–	–	–	–	–	–
Employees							
11	Chor 1999 (37)						
	Center branch	–	–	0	–	0	–
	Internal services unit	–	–	0	–	0	–
	General board	–	–	0	–	0	–
	Ilha do governador branch	–	–	0	–	0	–
16	Fonseca 2004 (40)	–	–	0.2	3.2	1.1	3.0
18	Giles 1991 (42)	–	–	2.8	–	1.0	–
36	Nakamura 1999 (59)	NA	NA	NA	NA	–0.1	0.7
38	Niedhammer 2000 (61)	–	–	0.4	(0.4, 0.4)	0.4	(0.5, 0.3)
60	Wada 2005 (78)	–	–	0.1	(0.1)	0	(0.1)
Elderly							
19	Gunnell 2000 (43)	–	–	2.1	(1.6, 2.7)	1.7	(1.3, 2.2)
27	Kuczmariski 2001 (22) 60+ year	–	–	2.7	0.1	2.5	0.1
59	van der Voort 2000 (77)	2.0	–	–	–	–	–
63	Zhang 1993 (86)	–	–	–	–	–	–

*Represents the 99% CI.

†Standard error.

‡Mean relative errors.

–, not reported; BMI, body mass index; NA, not applicable.

Only four provided data for the entire sample as well as according to the sex breakdowns (Nakamura (59) and Nawaz (60) were exceptions as these studies only examined women).

In the 18 studies that had mean error data for the overall sample, height was overestimated in all but one (39). Doll and Fairburn’s examination of healthy controls and individuals with bulimia nervosa found that while healthy controls overestimated their height, there was no significant difference between self-report and directly measured values for those with bulimia. The mean error for the rest of the studies ranged from a low of 0.6 cm (35,62) to a high of 7.5 cm (45). In those reports that had data on men, 29 studies reported that height was overestimated (range 0.1 cm–5.0 cm) and in two it was underestimated (32,44) (mean error of –1.3 cm and –0.7 cm respectively). Two studies (28,37) had mixed results. LeJarraga (28) examined

parents of children attending growth clinics and found that parents in private clinics overestimated their height, while those from public clinics tended to slightly underestimate their height. Chor’s (37) study of bank employees found no difference between the self-reported and measured values. Most studies also found that height was overestimated in women except in four studies where height was underestimated (32,44,59,65) and one that had mixed results (67). Standard deviations of the differences between reported and actual measures were large ranging from 0.1 cm in both men and women (22) to 8.8 cm in men and to 11.2 cm (67) in women. Studies that had data available for the overall sample were also plotted on a forest plot (see Fig. 2). The studies are ordered by sample size [smallest to largest with the three studies that had long time lags (34,45,72) (mean of over 1 month) plotted at the top of the figure].

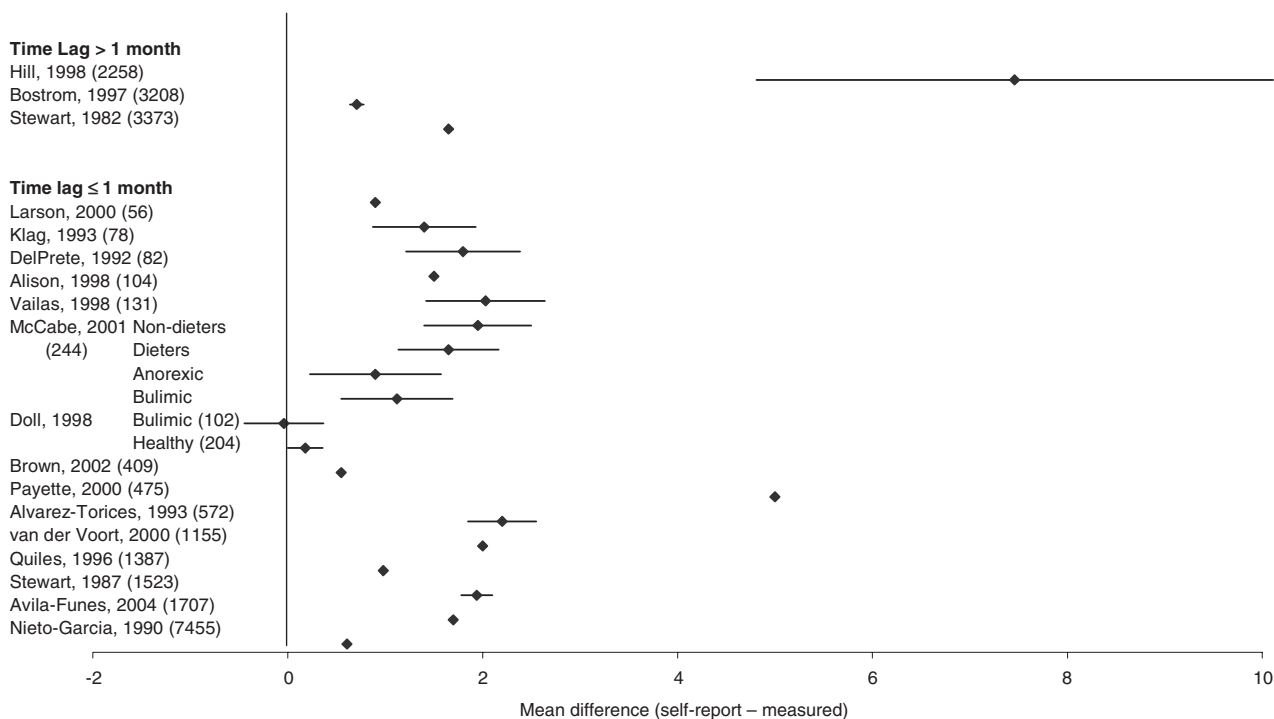


Figure 2 Mean differences in height for studies with available data on the total sample, in ascending order by sample size.

Weight

Table 5 displays the mean differences for weight for 56 studies that provided data on self-reported and measured weight. With weight as an outcome, 14 studies provided mean differences and associated variability measures for the overall sample while 24 provided them for either men or women as subgroups.

In examining the overall results, the tendency is for weight to be underestimated except in two studies (in Avila-Funes's work (31) in a Mexican population and in the anorexic sample of McCabe's study (57) where weight was overestimated). Similarly in studies reporting data separately for women, all but three found that women underestimated their weight when compared with the directly measured values (-0.1 kg (58) to -6.5 kg (60) in a study of obese women with BMIs between 35 and 40 kg m⁻²); Avila-Funes (31) study of Mexican women, and the subgroup of women in Ziebland's study (80) who had BMIs less than 20 reported overestimating their weight by approximately 0.7 kg and 0.8 kg respectively, and Wada's analysis (78) of Japanese women found no difference between the reported and measured values. The trend for men was also in the direction of under-reporting with 27 of the 34 studies that calculated mean differences for men finding that men under-reported their weight (range -0.1 kg to -3.2 kg (80)). The range in standard deviations was slightly smaller in weight than in height from 0.1 kg

in a large national American survey (22) to 4.5 kg in a smaller study of veterans (49) in men and in women ranged from 0.1 kg again in the national survey (22) to 10.1 kg in a small study of 96 obese women in the USA (60). These values represent a significant deviation from their measured values for some individuals. Figure 3 depicts a forest plot of the studies that had data available for the overall sample; the studies with longer time lags are plotted at the top.

Body mass index

Self-reported BMI is based on self-reports of both height and weight and measured BMI is based on measured values of both of these variables. BMI values from Masheb and Grilo (56) were not included since both their self-report and measured BMIs were calculated from the same directly measured height variable (i.e. self-reported height was not collected). Twenty-nine studies are included in Table 6, 18 reported mean difference and variability estimates (either by sex or overall) to compare self-reported BMI with the directly measured values. Eleven studies are included in the forest plot (Fig. 4). All but two studies found that the tendency was for BMI to be underestimated or accurately reported relative to the measured values. The range for men was from no difference (78) to -2.1 kg m⁻² (62) and for women ranged from no difference (78) to -2.2 kg m⁻² in Nawaz *et al.*'s sample (60) of

Table 5 Mean differences in weight (self-report–direct measure)

ID	Study – first author (reference)	Mean difference (kg)					
		Total	SD (95% CI)	Men	SD (95% CI)	Women	SD (95% CI)
General population							
2	Alvarez-Torices 1993 (30)	-0.6	(-0.9, -0.3)	-0.2	(-0.5, +0.1)	-0.9	(-1.4, -0.5)
3	Avila-Funes 2004 (31)	0.6	-	0.4	-	0.7	-
5	Bolton-Smith 2000 (32)	-	-	-0.6	3.5	-1.0	2.6
6	Booth 2000 (33)	-	-	-	-	-	-
7	Bostrom 1997 (34)	-1.2	(-1.4, -1.1)	-0.7	-	-1.6	-
24	Jalkanen 1987 (48)	-	-	-0.4	3.0	-0.6	2.0
27	Kuczmarski 2001 (22) 20–59 years	-	-	0.4	0.1	-1.5	0.1
29	Kuskowska-Wolk 1992 (51)	-	-	-1.0	-	-1.5	-
30	Lackland 1990 (53)	-	-	-	-	-	-
31	Larson 2000 (54)	-	-	-0.5	-	-	-
44	Quiles 1996 (82)	-1.1	-	-1.0	-	-1.2	-
45	Roberts 1995 (66)	-	-	-0.2	2.8	-1.1	2.6
47	Rowland 1990 (17)	-	-	0.4	3.0	-1.0	3.0
50	Schmidt 1993 (69)	-0.1	3.2	0.3	-	-0.3	-
52	Spencer 2002 (71)	-	-	-1.9	2.9	-1.4	2.5
54	Stewart 1987 (73)	-0.6	(-0.8, -0.4)*	-	-	-	-
55	Stunkard 1981 (26)	-	-	-	-	-	-
	US site	-1.2	-	-	-	-	-
	Danish site < 40	-	-	-1.1	-	-1.6	-
	Danish site ≥ 40	-	-	-1.3	-	-2.5	-
64	Ziebland 1996 (80)	-	-	-	-	-	-
	BMI < 20	-	-	2.6	-	0.8	-
	BMI 20–24	-	-	-0.1	-	-0.5	(-0.7, -0.2)
	BMI 25–29	-	-	-1.2	(-1.6, -0.9)	-1.8	(-2.2, -1.5)
	BMI > 30	-	-	-3.2	(-4.3, -2)	-3.3	(-4.2, -2.4)
Overweight or weight loss participants							
1	Allison 1998 (29)	-1.4	-	-3.0	-	-0.9	-
14	DelPrete 1992 (38)	-2.3	1.9	-2.1	1.5	-2.4	1.9
37	Nawaz 2001 (60)	-	-	-	-	-	-
	25 ≤ BMI < 30	NA	NA	NA	NA	-0.3	4.6
	30 ≤ BMI ≤ 35	NA	NA	NA	NA	-1.6	5.7
	35 < BMI ≤ 40	NA	NA	NA	NA	-6.5	0.2
	BMI > 40	NA	NA	NA	NA	-5.2	10.1
56	Tell 1987 (74)	-2.7	3.2	-2.8	3.4	-2.3	2.5
Elderly							
19	Gunnell 2000 (43)	-	-	-1.9	(-2.7, -1.2)	-1.2	(-1.7, -0.8)
27	Kuczmarski 2001 (22) 60+ years	-	-	0.5	0.1	-0.6	0.1
32	Lawlor 2002 (55)	NA	NA	NA	NA	-1.0	(-1.1, -0.8)
63	Zhang 1993 (86)	-	-	-	-	-	-
Eating disorders							
15	Doll 1998 (39)	-	-	-	-	-	-
	Healthy controls	-1.0	(-1.3, -0.7)	-	-	-	-
	Bulimia nervosa	-0.3	(-0.9, -0.2)	-	-	-	-
34	Masheb 2001 (56)	-0.9	-	-	-	-	-
35	McCabe 2001 (57)	-	-	-	-	-	-
	Anorexics	0.3	1.9	-	-	-	-
	Bulimics	-0.6	2.0	-	-	-	-
	Dieters	-3.5	2.9	-	-	-	-
	Non-dieters	-1.1	2.9	-	-	-	-
9	Brunet 2003 (87)	NA	NA	NA	NA	-1.1	1.8
10	Cash 1989 (36)	NA	NA	NA	NA	-	-
12	Clemente 2004 (81)	-	-	-0.3	3.0	-0.2	2.2
22	Imrhan 1996 (46)	-	-	-	-	-	-
23	Jacobson 2001 (47)	-	-	0.5	-	-1.9	-
51	Smith 1992 (70)	-1.6	3.7	-	-	-	-
62	Wing 1979 (79)	-	-	-	-	-	-

Table 5 Continued

ID	Study – first author (reference)	Mean difference (kg)					
		Total	SD (95% CI)	Men	SD (95% CI)	Women	SD (95% CI)
Employees							
11	Chor 1999 (37)						
	Center branch	–	–	–1.1	–	–1.0	–
	Internal services unit	–	–	–1.3	–	–0.7	–
	General board	–	–	–0.7	–	–0.8	–
	Ilha do governador branch	–	–	–1.1	–	–0.7	–
16	Fonseca 2004 (40)	–	–	–1.0	3.5	–1.1	3.0
17	Forster 1988 (41)	–2.3	–	–	–	–	–
36	Nakamura 1999 (59)	NA	NA	NA	NA	–0.2	1.8
38	Niedhammer 2000 (61)	–	–	–0.5	(–0.6, –0.5)	–0.9	(–1.0, –0.7)
60	Wada 2005 (78)	–	–	0	(–0.1, 0)	0	(–0.1, 0.2)
Parents/families							
53	Stewart 1982 (72)	–1.1	5.5	–	–	–	–
57	Tienboon 1992 (75)	–	–	–0.3	3.0	–1.2	2.5
Special populations							
13	de Araujo 2003 (58)	–	–	–0.9	–	–0.1	–
21	Hill 1998 (45)	–0.9	(–2.0, 0.3)	–	–	–	–
25	Kinney 1988 (49)	NA	NA	–3.0	4.5	NA	NA
26	Klag 1993 (50)	–1.5	0.3 [†]	–	–	–	–
28	Kuskowska-Wolk 1989 (52)	–	–	–0.5	–	–0.6	–
39	Nieto-Garcia 1990 (62) [‡]	–0.7	(–0.8, –0.6)	–0.3	(–0.5, –0.2)	–1.0	(–1.1, –0.9)
41	Palta 1982 (63)	–	–	–1.5	3.2	–2.4	4.0
42	Payette 2000 (64)	–0.2	–	0.2	–	–0.3	–
43	Pirie 1981 (65)	–	–	–0.5	3.3	–1.9	3.0
48	Santillan 2003 (68)	–	–	–	–	–	–
58	Vailas 1998 (76)	–0.5	3.1	–0.3	3.6	–0.7	2.9
Experimental design							
4	Black 1998 (85)						
	Informed group	–0.9	–	–	–	–	–
	Uninformed group	–1.3	–	–	–	–	–
40	O'Connell 2005 (84)	–2.0	–	–	–	–	–

*Represents the 99% CI.

[†]Standard error.[‡]Mean relative errors.

–, not reported; BMI, body mass index; NA, not applicable.

women who had BMI values over 40. Standard deviations were smaller than for the individual values of height and weight ranging from 0 to 1.6 kg m^{–2} for men and from 0 to 2.2 kg m^{–2} for women. Bolton-Smith's work (32), which sampled the general population aged 25–64 years in Scotland from general practitioner's registries, had results that were inconsistent with this trend, finding instead that BMI tended to be overestimated by a mean difference 0.2 ± 1.4 kg m^{–2} in men and 0.2 ± 1.3 kg m^{–2} in women. Quiles and Vioque's (82) general population sample of residents 15 and over also found BMI to be overestimated in men by 0.5 kg m^{–2} and in women by 0.9 kg m^{–2} (standard deviations not provided).

Quantitative analysis

Quantitative analysis is possible when the data are sufficiently consistent across studies in terms of statistical, clinical and methodological characteristics. Tables 4–6 have loosely grouped the studies according to their main populations of interest, and the quality assessment would permit a further subdivision based on study quality. Because of the substantial clinical and methodological inconsistency across the study reports and the large amount of missing data (i.e. mean differences and corresponding measures of variance), we did not combine the data quantitatively. Although all studies should be reporting mean differences

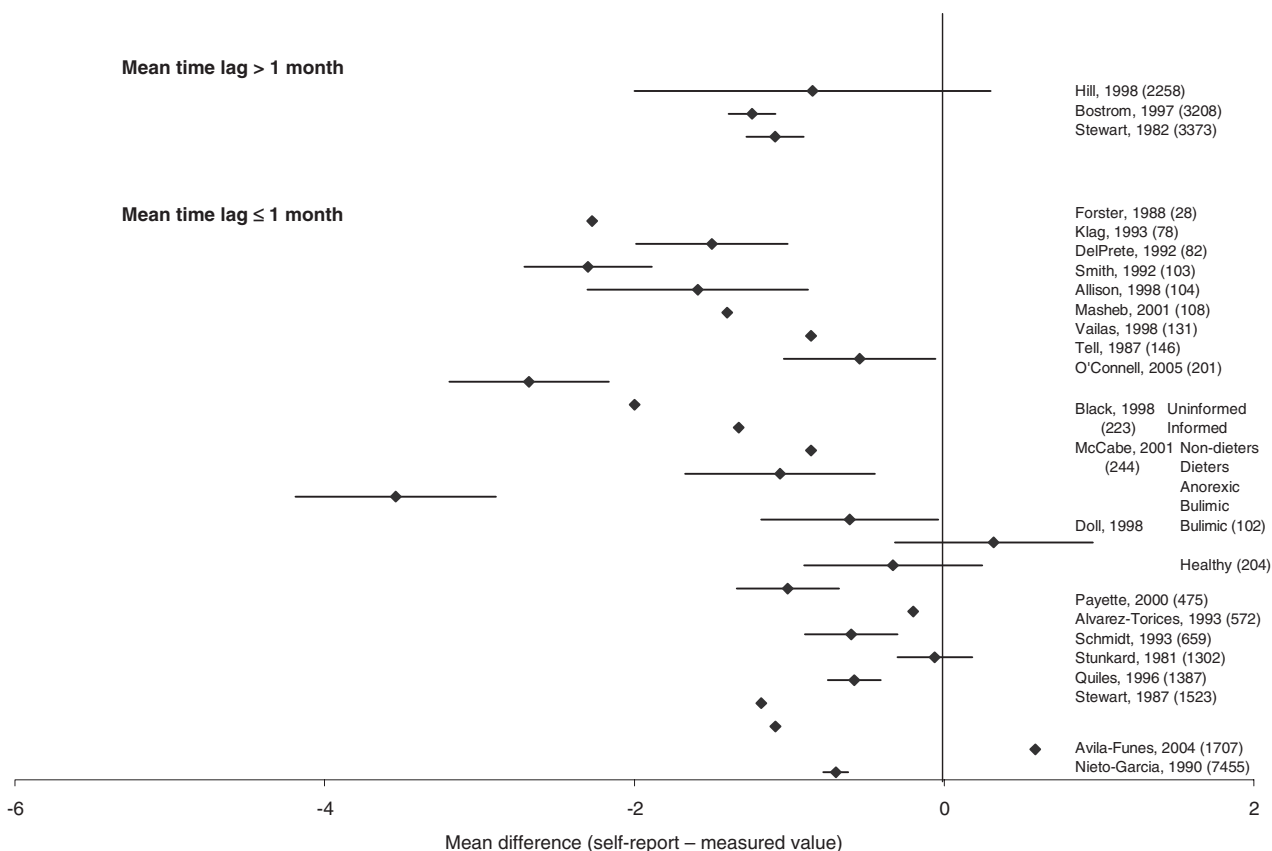


Figure 3 Mean differences in weight for studies with available data on the total sample, in ascending order by sample size.

and variance estimates, some authors only report proportions or correlation coefficients, which creates difficulties for pooling the estimates. A small subset of authors was contacted in an attempt to acquire missing data, but no data were received to cover the information gaps.

Discussion

The trend in studies where data were available was for height to be overestimated and weight and BMI to be underestimated. This trend was evident for both women and men. The size of the mean errors in weight varied depending on the study and the population being examined. In many cases, they were relatively small but in some instances, such as Ziebland’s study (80) of individuals with BMIs over 30 and Nawaz’s (60) work with populations with BMIs over 35, the mean differences in weight were quite high (underestimated by 3–6 kg). In many of the studies, the standard deviations were large indicating that there is a great deal of individual variability in reporting of results for all of the outcomes examined. This can have important implications for clinical and population health practice as even a small difference in body weight has the potential to substantially modify the BMI classification,

which could mean that current obesity prevalence estimates are underestimated.

The findings in this review are similar to those of Engstrom *et al.* (21) who reviewed the accuracy of height and weight in women and adolescent females. They examined 26 studies on the accuracy of self-reported height and found that in 21 of them height was overestimated. Similarly, they found weight to be underestimated in all of the 34 studies that they reviewed on the accuracy of self-reported weight. In addition, Bowman and Delucia (88) conducted a meta-analysis of the accuracy of self-reported weight and found that there was significant bias in the self-reported measures, with a tendency for self-reported weights to be underestimated. In spite of the trend in the present review of weight and BMI being underestimated and height being overestimated, there were too many gaps in the data to undertake a quantitative analysis or to get a comprehensive understanding of the relationship between self-report and direct measures.

Although studies are for the most part of good quality, Tables 4–6 highlight the degree to which there are gaps in reporting; studies lack vital information that allow conclusions to be verified and analyses replicated. Studies in this review have been ongoing since 1979 and would have

Table 6 Mean differences in BMI (self-report–direct measure)

ID	Study – first author (reference)	Mean difference (kg m ⁻²)					
		Total	SD (95% CI)	Men	SD (95% CI)	Women	SD (95% CI)
General population							
2	Alvarez-Torices 1993 (30)	-1.0	(-1.2, -0.8)	-0.8	(-1.0, -0.6)	-1.2	(-1.5, -0.9)
3	Avila-Funes 2004 (31)	-0.4	-	-0.3	-	-0.5	-
5	Bolton-Smith 2000 (32)	-	-	0.2	1.4	0.2	1.3
6	Booth 2000 (33)	-	-	-	-	-	-
7	Bostrom 1997 (34)	-0.7	(-0.7, -0.6)	-0.4	-	-0.9	-
27	Kuczmarski 2001 (22) 20–59 years	-	-	-0.3	0	0	0
29	Kuskowska-Wolk 1992 (51)	-	-	-0.4	-	-0.8	-
30	Lackland 1990 (53)	-	-	-	-	-	-
44	Quiles 1996 (82)	-0.7	-	0.5	-	0.9	-
52	Spencer 2002 (71)	-	-	-1.0	1.2	-0.7	1.3
54	Stewart 1987 (73)	-0.8	(-0.8, -0.6)*	-	-	-	-
Overweight or weight loss participants							
1	Allison 1998 (29)	-1.6	-	-2.0	-	-1.4	-
37	Nawaz 2001 (60)						
	25 ≤ BMI < 30	NA	NA	NA	NA	-0.2	0.7
	30 ≤ BMI ≤ 35	NA	NA	NA	NA	-0.3	1.3
	35 < BMI ≤ 40	NA	NA	NA	NA	-1.5	1.9
	BMI > 40	NA	NA	NA	NA	-2.2	2.2
Students							
12	Clemente 2004 (81)	-	-	-0.4	1.1	-0.8	1.1
23	Jacobson 2001 (47)	-	-	-0.2	-	-0.6	-
Employees							
11	Chor 1999 (37)						
	Center branch	-	-	-0.2	-	-0.3	-
	Internal services unit	-	-	-0.6	-	-0.2	-
	General board	-	-	-0.2	-	-0.3	-
	Ilha do governador branch	-	-	-0.4	-	-0.3	-
16	Fonseca 2004 (40)	-	-	-0.4	1.6	-0.8	1.6
36	Nakamura 1999 (59)	NA	NA	NA	NA	-0.1	0.8
38	Niedhammer 2000 (61)	-	-	-0.3	(-0.3, -0.3)	-0.4	(-0.4, -0.5)
60	Wada 2005 (78)	-	-	0	(-0.1, 0)	0	(-0.1, 0.1)
Elderly							
19	Gunnell 2000 (43)	-	-	-1.3	(-1.6, -1.0)	-1.1	(-1.3, -0.9)
27	Kuczmarski 2001 (22) 60+ years	-	-	-0.8	0	-0.8	0.0
Special populations							
13	de Araujo 2003 (58)	-	-	-	-	-	-
15	Doll 1998 (39)						
	Healthy controls	-0.4	(-0.6, -0.3)	-	-	-	-
	Bulimia nervosa	-0.2	(-0.4, 0.1)	-	-	-	-
21	Hill 1998 (45)	-1.3	(-1.3, -1.2)	-	-	-	-
26	Klag 1993 (50)	-0.8	0.1 [†]	-	-	-	-
28	Kuskowska-Wolk 1989 (52)	-	-	-0.4	-	-0.7	-
39	Nieto-Garcia 1990 (62) [‡]	-1.8	(-1.9, -1.7)	-2.1	(-2.3, -2.0)	-1.6	(-1.7, -1.4)
48	Santillan 2003 (68)	-	-	-	-	-	-
57	Tienboon 1992 (75)	-	-	-1.0	1.2	-1.0	1.2

*Represents the 99% CI.

[†]Standard error.[‡]Mean relative errors.

-, not reported; BMI, body mass index; NA, not applicable.

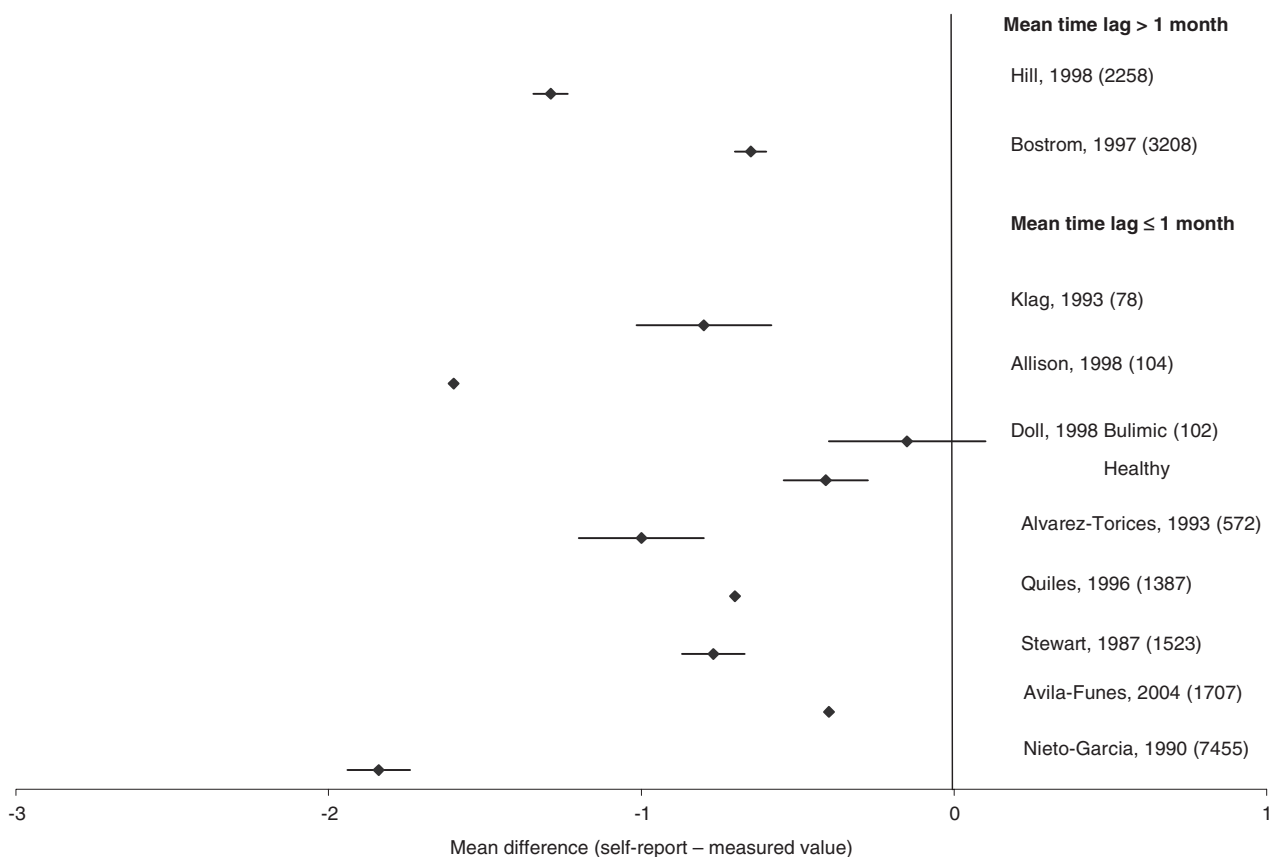


Figure 4 Mean differences in BMI for studies with available data on the total sample, in ascending order by sample size.

benefited from the development of some standard reporting criteria to address some of these gaps over this 26-year period. In 1988, Engstrom (89) published an article outlining the statistics and procedures that should be reported whenever assessing the reliability of physical measures such as weight and height. She suggested that values such as the mean, minimum and maximum differences, standard deviations of the net differences, technical error of measurement and clinically meaningful indices of agreement be included in all reports. However, as evidenced in the current review almost 20 years later, such data are still not routinely being reported. In addition, the authors of this current review recommend that the elapsed time between self-report and direct measures should also be reported, as changes in variables such as weight can occur over short periods of time and the length of time between measures should be controlled in any analyses that are undertaken. The order of measurement should also be noted as there is evidence to indicate that reporting of weight in women is more accurate when it occurs after actual weight measurement (36). Measurement conditions such as who is taking the measures, what equipment is being used, what clothing is worn for the weighing as well as what instructions are given when estimating self-report weights (e.g. is this self-reported weight estimated with or without clothes, in the

morning or midday and after eating or fasting?) should also be provided.

Adhering to consistent reporting criteria would increase the comparability of results and enable us to further understand the relationship between the measures. With more complete data, it may be possible to develop correction factors that could be applied to self-reported data when direct measurement is not feasible. If data are adequately reported by subgroups (e.g. sex, age, education, ethnicity and immigration status), it may also be possible to make adjustments based on these characteristics, which may be important as studies demonstrate that demographic characteristics can have an influence on the degree of reporting error. For instance, research has shown that self-reports from older populations are less reliable (22), that unemployed, retired or disabled women are more likely to under-report their BMIs than employed women (60), and that men are more likely to over-report their height than women (38). Understanding the characteristics of inaccurate reporters is particularly important given the large values of some of the standard deviations, which implies that for some individuals the estimates are extremely inaccurate, which may bias average results and have important implications for health planning for these populations.

Conclusions

To the authors' knowledge, this review represents the most comprehensive attempt to examine the relationship between self-reported and directly measured height, weight and BMI in both men and women. Overall, the data show trends of underestimating weight and BMI and overestimating height, although the degree of the trend varies for men and women, and between studies; the variability in the estimates is high and no overall effect size could be estimated. Rising obesity rates are an increasing concern for population health and it is therefore essential that the estimates upon which decisions are being made are as accurate as possible. This should be addressed by improving the quality of reporting of future studies in this area, making recommendations based on more comprehensive data about whether (and in what circumstances) self-report measures can be used and then developing correction factors to increase the accuracy of reporting in situations where direct measurement is not possible.

Conflict of Interest Statement

No conflict of interest was declared.

Acknowledgement

Dr Moher is supported, in part, by a University of Ottawa Research Chair.

References

1. Health Canada. *Body Mass Index (BMI) Nomogram*. 19 September 2003. [WWW document]. URL http://www.hc-sc.gc.ca/fn-an/nutrition/weights-poids/guide-ld-adult/bmi_chart_java_graph_imc_java_e.html (accessed 15 April 2006).
2. Katzmarzyk PT. The Canadian obesity epidemic, 1985–1998. *CMAJ* 2002; **166**: 1039–1040.
3. World Health Organization. *Obesity and Overweight*. [WWW document]. URL <http://www.who.int/dietphysicalactivity/publications/facts/obesity/en/prinwt.html> (accessed 10 April 2006).
4. World Health Organization. *World Health Report 2002 Reducing Risks, Promoting Healthy Life*. [WWW document]. URL <http://www.who.int/whr/2002/en/> (accessed 10 April 2006).
5. Tjepkema M. *Measured Obesity. Adult Obesity in Canada: Measured Height and Weight. Nutrition: Findings from the Canadian Community Health Survey*. Statistics Canada: Ottawa, 2005. Statistics Canada catalogue number 82-620-MWE2005001.
6. Shields M. *Measured Obesity. Overweight Canadian Children and Adolescents. Nutrition: Findings from the Canadian Community Health Survey*. Statistics Canada: Ottawa, 2005. Statistics Canada catalogue 82-620-MWE2005001.
7. Katzmarzyk PT, Janssen I. The economic costs associated with physical inactivity and obesity in Canada: an update. *Can J Appl Physiol* 2004; **29**: 90–115.
8. Larsson D, Hemmingsson T, Allebeck P, Lundberg I. Self-rated health and mortality among young men: what is the relation and how may it be explained? *Scand J Public Health* 2002; **30**: 259–266.
9. Af Sillen U, Nilsson J, Mansson N, Nilsson PM. Self-rated health in relation to age and gender: influence on mortality risk in the Malmo Preventive Project. *Scand J Public Health* 2005; **33**: 183–189.
10. Heistaro S, Jousilahti P, Lahelma E, Vartiainen E, Puska P. Self rated health and mortality: a long term prospective study in eastern Finland. *J Epidemiol Community Health* 2001; **55**: 227–232.
11. Singleton RA, Straits BC, Miller Straits M. *Approaches to Social Research*, 2nd edn. Oxford University Press: New York, 1993.
12. Adlaf EM, Mates D. *A Comparison of Telephone and Personal Interview Surveys for Estimating Alcohol and Other Drug Use*. Addiction Research Foundation: Toronto, 1994.
13. Midanik LT, Greenfield TK, Rogers JD. Reports of alcohol-related harm: telephone versus face-to-face interviews. *J Stud Alcohol* 2001 January; **62**: 74–78.
14. Wiener L, Riekert K, Ryder C, Wood LV. Assessing medication adherence in adolescents with HIV when electronic monitoring is not feasible. *AIDS Patient Care Stds* 2004; **18**: 527–538.
15. Statistics Canada. *Canadian Community Health Survey (CCHS): Cycle 1.1: Extending the Wealth of Health Data in Canada*. 2003. [WWW document]. URL <http://www.statcan.ca/english/concepts/health/cchsinfo.htm> (accessed 10 February 2006).
16. St-Pierre M, Béland Y. *Mode effects in the Canadian Community Health Survey: A comparison of CAPI and CATI. Proceedings of the American Statistical Association Meeting: Survey Research Methods*. American Statistical Association: Toronto, 2004.
17. Rowland ML. Self-reported weight and height. *Am J Clin Nutr* 1990; **52**: 1125–1133.
18. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17**: 1–12.
19. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; **52**: 377–384.
20. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, Petticrew M, Altman DG. Evaluating non randomized intervention studies. *Health Technol Assess* 2003; **7**: 1–173.
21. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. *J Midwifery Womens Health* 2003; **48**: 338–345.
22. Kuczmarski MF, Kuczmarski RJ, Najjar M. Effects of age on validity of self-reported height, weight, and body mass index: findings from the Third National Health and Nutrition Examination Survey, 1988–1994. *J Am Diet Assoc* 2001; **101**: 28–34.
23. Villanueva EV. The validity of self-reported weight in US adults: a population based cross-sectional study. *BMC Public Health* 2001; **1**: 11. Epub 2001 Nov 6.
24. Gillum RF, Sempos 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.
25. US Department of Health and Human Services, Centers for Disease Control and Prevention. National Centre for Health Statistics. *National Health and Nutrition Survey 2005–2006 Overview*. 2005. [WWW document]. URL http://www.cdc.gov/nchs/data/nhanes/OverviewBrochureEnglish_May05.pdf (accessed 15 April 2006).
26. Stunkard AJ, Albaum JM. The accuracy of self-reported weights. *Am J Clin Nutr* 1981; **34**: 1593–1599.
27. Schlichting P, Hoilund-Carlsen PF, Quaade F. Comparison of self-reported height and weight with controlled height and weight in women and men. *Int J Obes* 1981; **5**: 67–76.

28. LeJarraga H, Laspair M, Adamo P. Validity of reported parental height in outpatient growth clinics in Buenos Aires city. *Ann Hum Biol* 1995; 22: 163–166.
29. Allison DB, Nathan JS, Albu JB, Heymsfield SB, Duprat LJ, Pi-Sunyer FX. Measurement challenges and other practical concerns when studying massively obese individuals. *Int J Eat Disord* 1998; 24: 275–284.
30. Alvarez-Torices JC, Franch-Nadal J, Alvarez-Guisasola F, Hernandez-Mejia R, Cueto-Espinar A. Self-reported height and weight and prevalence of obesity. Study in a Spanish population. *Int J Obes Relat Metab Disord* 1993; 17: 663–667.
31. Avila-Funes JA, Gutierrez-Robledo LM, Ponce De Leon RS. Validity of height and weight self-report in Mexican adults: results from the national health and aging study. *J Nutr Health Aging* 2004; 8: 355–361.
32. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self-reported height and weight in an adult Scottish population. *J Epidemiol Community Health* 2000; 54: 143–148.
33. Booth ML, Hunter C, Gore CJ, Bauman A, Owen N. The relationship between body mass index and waist circumference: implications for estimates of the population prevalence of overweight. *Int J Obes Relat Metab Disord* 2000; 24: 1058–1061.
34. Bostrom G, Diderichsen F. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *Int J Epidemiol* 1997; 26: 860–866.
35. Brown JK, Feng JY, Knapp TR. Is self-reported height or arm span a more accurate alternative measure of height? *Clin Nurs Res* 2002; 11: 417–432.
36. Cash TF, Counts B, Hangen J, Huffine CE. How much do you weigh?: determinants of validity of self-reported body weight. *Percept Mot Skills* 1989; 69: 248–250.
37. Chor D, Coutinho ES, Laurenti R. Reliability of self-reported weight and height among state bank employees. *Rev Saude Publica* 1999; 33: 16–23.
38. DelPrete LR, Caldwell M, English C, Banspach SW, Lefebvre C. Self-reported and measured weights and heights of participants in community-based weight loss programs. *J Am Diet Assoc* 1992; 92: 1483–1486.
39. Doll HA, Fairburn CG. Heightened accuracy of self-reported weight in bulimia nervosa. A useful cognitive ‘distortion’. *Int J Eat Disord* 1998; 24: 267–273.
40. Fonseca MJ, Faerstein E, Chor D, Lopes CS. [Validity of self-reported weight and height and the body mass index within the ‘Pro-saude’ study]. [Portuguese]. *Rev Saude Publica* 2004; 38: 392–398.
41. Forster JL, Jeffery RW, Snell MK. One year follow up study to a worksite weight control program. *Prev Med* 1988; 17: 129–133.
42. Giles E, Hutchison DL. Stature- and age-related bias in self-reported stature. *J Forensic Sci* 1991; 36: 765–780.
43. Gunnell D, Berney L, Holland P, Maynard M, Blane D, Frankel S, Smith GD. How accurately are height, weight and leg length reported by the elderly, and how closely are they related to measurements recorded in childhood? *Int J Epidemiol* 2000; 29: 456–464.
44. Hensley WE. The measurement of height. *Adolescence* 1998; 33: 629–635.
45. Hill A, Roberts J. Body mass index: a comparison between self-reported and measured height and weight. *J Public Health Med* 1998; 20: 206–210.
46. Imrhan SN, Imrhan V, Hart C. Can self-estimates of body weight and height be used in place of measurements for college students. *Ergonomics* 1996; 39: 1445–1453.
47. Jacobson BH, DeBock DH. Comparison of Body Mass Index by self-reported versus measured height and weight. *Percept Mot Skills* 2001; 92: 128–132.
48. Jalkanen L, Tuomilehto J, Tanskanen A, Puska P. Accuracy of self-reported body weight compared to measured body weight. A population survey. *Scand J Soc Med* 1987; 15: 191–198.
49. Kinney EL. Accuracy of self-reported weight in a non-normal population. *Clin Invest Med* 1988; 11: 347–350.
50. Klag MJ, He J, Mead LA, Ford DE, Pearson TA, Levine DM. Validity of physicians’ self reports of cardiovascular disease risk factors. *Ann Epidemiol* 1993; 3: 442–447.
51. Kuskowska-Wolk A, Bergstrom R, Bostrom G. Relationship between questionnaire data and medical records of height, weight and body mass index. *Int J Obes Relat Metab Disord* 1992; 16: 1–9.
52. Kuskowska-Wolk A, Rossner S. The ‘true’ prevalence of obesity. A comparison of objective weight and height measures versus self-reported and calibrated data. *Scand J Prim Health Care* 1989; 7: 79–82.
53. Lackland DT, Wheeler FC. The need for accurate nutrition survey methodology: the South Carolina experience. *J Nutr* 1990; 120: 1433–1436.
54. Larson MR. Social desirability and self-reported weight and height. *Int J Obes Relat Metab Disord* 2000; 24: 663–665.
55. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Agreement between measured and self-reported weight in older women. Results from the British Women’s Heart and Health Study. *Age Ageing* 2002; 31: 169–174.
56. Masheb RM, Grilo CM. Accuracy of self-reported weight in patients with binge eating disorder. *Int J Eat Disord* 2001; 29: 29–36.
57. McCabe RE, McFarlane T, Polivy J, Olmstead MP. Eating disorders dieting and the accuracy of self-reported weight. *Int J Eat Disord* 2001; 29: 59–64.
58. de Araujo DS, de Araujo CG. Self-perception and dissatisfaction with weight does not depend on the frequency of physical activity. *Arq Bras Cardiol* 2003; 80: 235–249.
59. Nakamura K, Hoshino Y, Kodama K, Yamamoto M. Reliability of self-reported body height and weight of adult Japanese women. *J Biosoc Sci* 1999; 31: 555–558.
60. Nawaz H, Chan W, Abdulrahman M, Larson D, Katz DL. Self-reported weight and height: implications for obesity research. *Am J Prev Med* 2001; 20: 294–298.
61. Niedhammer I, Bugel I, Bonenfant S, Goldberg M, Leclerc A. Validity of self-reported weight and height in the French GAZEL cohort. *Int J Obes Relat Metab Disord* 2000; 24: 1111–1118.
62. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. *Epidemiology* 1990; 1: 146–152.
63. Palta M, Prineas RJ, Berman R, Hannan P. Comparison of self-reported and measured height and weight. *Am J Epidemiol* 1982; 115: 223–230.
64. Payette H, Kergoat MJ, Shatenstein B, Boutier V, Nadon S. Validity of self-reported height and weight estimates in cognitively-intact and impaired elderly individuals. *J Nutr Health Aging* 2000; 4: 223–228.
65. Pirie P, Jacobs D, Jeffery R, Hannan P. Distortion in self-reported height and weight data. *J Am Diet Assoc* 1981; 78: 601–606.
66. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? *Public Health* 1995; 109: 275–284.
67. Rona RJ, Chinn S, Manning R. The validity of reported parental height in inner city areas in England. *Ann Hum Biol* 1989; 16: 41–44.
68. Santillan AA, Camargo CA. Body mass index and asthma among Mexican adults: the effect of using self-reported vs mea-

- sured weight and height. *Int J Obes Relat Metab Disord* 2003; 27: 1430–1433.
69. Schmidt MI, Duncan BB, Tavares M, Polanczyk CA, Pellanda L, Zimmer PM. Validity of self-reported weight – a study of urban Brazilian adults. *Rev Saude Publica* 1993; 27: 271–276.
70. Smith GT, Hohlsteing LA, Atlas JG. Accuracy of self-reported weight: covariation with binger or restrainer status and eating disorder symptomatology. *Addict Behav* 1992; 17: 1–8.
71. Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr* 2002; 5: 561–565.
72. Stewart AL. The reliability and validity of self-reported weight and height. *J Chronic Dis* 1982; 35: 295–309.
73. Stewart AW, Jackson RT, Ford MA, Beaglehole R. Underestimation of relative weight by use of self-reported height and weight. *Am J Epidemiol* 1987; 125: 122–126.
74. Tell GS, Jeffery RW, Kramer FM, Snell MK. Can self-reported body weight be used to evaluate long-term follow-up of a weight-loss program? *J Am Diet Assoc* 1987; 87: 1198–1201.
75. Tienboon P, Wahlqvist ML, Rutishauser IH. Self-reported weight and height in adolescents and their parents. *J Adolesc Health* 1992; 13: 528–532.
76. Vilais LI, Nitzke SA. Self-reported versus measured weight and height in an older adult meal program population. *J Gerontol A Biol Sci Med Sci* 1998; 53: M481–M483.
77. van der Voort DJ, Brandon S, Dinant GJ, van Wersch JW. Screening for osteoporosis using easily obtainable biometrical data. Diagnostic accuracy of measured, self-reported and recalled BMI, related costs of bone mineral density measurements. *Osteoporos Int* 2000; 11: 233–239.
78. Wada K, Tamakoshi K, Tsunekawa T, Otsuka R, Zhang H, Murata C *et al.* Validity of self-reported height and weight in a Japanese workplace population. *Int J Obes* 2005; 29: 1093–1099.
79. Wing RR, Epstein LH, Ossip DJ, LaPorte RE. Reliability and validity of self-report and observers' estimates of relative weight. *Addict Behav* 1979; 4: 133–140.
80. Ziebland S, Thorogood M, Fuller A, Muir J. Desire for the body normal: body image and discrepancies between self-reported and measured height and weight in a British population. *J Epidemiol Community Health* 1996; 50: 105–106.
81. Clemente L, Moreira P, Oliveira B, Almeida MD. Body mass index: sensitivity and specificity. *Acta Med Port* 2004; 17: 353–358. Epub 2004 December 20.
82. Quiles IJ, Vioque J. [Validity of notified anthropometric data for determining the prevalence of obesity]. [Spanish]. *Med Clinica* 1996; 106: 725–729.
83. Wang DY, DeStavola BL, Allen DS, Fentiman IS, Bulbrook RD, Hayward JL, Reed MJ. Breast cancer risk is positively associated with height. *Breast Cancer Res Treat* 1997; 43: 123–128.
84. O'Connell KJ, Osborne LM, Westhoff C. Measured and reported weight change for women using a vaginal contraceptive ring vs. a low-dose oral contraceptive. *Contraception* 2005; 72: 323–327.
85. Black DR, Taylor AM, Coster DC. Accuracy of self-reported body weight: stepped approach to model component assessment. *Health Educ Res* 1998; 13: 301–307.
86. Zhang J, Feldblum PJ, Fortney JA. The validity of self-reported height and weight in perimenopausal women. *Am J Public Health* 1993; 83: 1052–1053.
87. Brunet MEI. Development and validation of the Screening Test for at Risk Individuals for Eating Disorders (STARVED). *Dissertation Abstracts International Section A: Humanities and Social Sciences* 2003 March; 63(8A): 2804.
88. Bowman RL, DeLucia JL. Accuracy of self-reported weight: a meta analysis. *Behav Ther* 1992; 23: 637–655.
89. Engstrom JL. Assesment of the reliability of physical measures. *Res Nurs Health* 1988; 11: 383–389.

Appendix I: Data Extraction Form

1. Information about reviewer
– Date: – Initials:

2. Information about the study
Authors:
Language:

3. Study design:

Experimental Observational

Specify type:

Published Yes No Can't Tell
Dissertation or thesis Yes No Can't Tell

4. Verification of study eligibility

- a. Design – observational or experimental:
 Yes No Can't Tell
- b. Participants – 18 and over:
 Yes No Can't Tell
- c. Direct comparison of objective and subjective measures
 Yes No Can't Tell
- d. Proxy reporting Yes No Can't Tell
- e. Outcomes
Self-reported weight (any collection method)
(e.g. telephone, in person) Yes No Can't Tell

Measured weight Yes No Can't Tell
(any type of scale)

Self-reported height (any collection method)
(e.g. telephone, in person)

Yes No Can't Tell

Measured height (any method) (e.g. tape measure, ruler)

Yes No Can't Tell

5. Study characteristics

a. Participants:

b. Exclusion criteria:

c. Sample size:

– Enrolled

– Analysed:

– Reasons for exclusion:

d. Variables (and definition):

– Method of weight measurement:

– Method of self-reported weight:

– Method of height measurement:

– Method of height self-report:

– Summary measures and method of calculation
(e.g. BMI, waist to hip ratio, skinfold)

e. Setting:

6. Results