



Published in final edited form as:

Int J Obes (Lond). 2020 June ; 44(6): 1311–1318. doi:10.1038/s41366-019-0499-5.

Use and reporting of Bland–Altman analyses in studies of self-reported versus measured weight and height

Katherine M. Flegal¹, Barry Graubard², John P. A. Ioannidis^{1,3,4,5,6}

¹Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, 1265 Welch Road, Mail Code 5411, Stanford, CA 94305-5411, USA

²Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

³Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA

⁴Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA

⁵Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA

⁶Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

Abstract

Background/Objectives—Bland–Altman methods for assessing the agreement between two measures are highly cited. However, these methods may often not be used to assess agreement, and when used, they are not always presented or interpreted correctly. Our objective was to evaluate the use and the quality of reporting of Bland–Altman analyses in studies that compare self-reported with measured weight and height.

Methods—We evaluated the use of Bland–Altman methods in 394 published articles that compared self-reported and measured weight and height data for adolescents or adults. Six reporting criteria were developed: assessment of the normality of the distribution of differences, a complete and correctly labeled Bland–Altman plot displaying the mean difference and limits of agreement (LOA), numerical values and confidence intervals, standard errors, or standard deviations for mean difference, numerical values of LOA, confidence intervals for LOA, and prespecified criteria for acceptable LOA.

Results—Only 72/394 (18%) studies comparing self-reported with measured weight and height or BMI used some form of Bland–Altman analyses. No study using Bland–Altman analyses satisfied more than four of the six criteria. Of the 72 studies, 64 gave mean differences along with

Katherine M. Flegal, kmflegal@stanford.edu.

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information The online version of this article (<https://doi.org/10.1038/s41366-019-0499-5>) contains supplementary material, which is available to authorized users.

confidence intervals or standard deviations, 55 provided complete Bland–Altman plots that were appropriately labeled and described, 37 provided numerical values for LOA, 4 reported that they examined the normality of the distribution of differences, 3 provided confidence intervals for LOA, and 3 had prespecified criteria for agreement.

Conclusions—Bland–Altman methods appear to be infrequently used in studies comparing measured with self-reported weight, height, or BMI, and key information is missing in many of those that do use Bland–Altman methods. Future directions would be defining acceptable LOA values and improving the reporting and application of Bland–Altman methods in studies of self-reported anthropometry.

Introduction

In 1986, Bland and Altman published an article [1] on methods of measurement comparison. Their purpose was to discuss methods of comparing two measures of the same thing, with the goal of assessing to the extent to which one method could be used in place of the other. This is the exact goal of many studies of self-reported height or weight compared with measured height and weight. According to the abstract: “in clinical measurement comparison of a new measurement technique with an established one is often needed to see whether they agree sufficiently for the new to replace the old. Such investigations are often analyzed inappropriately, notably by using correlation coefficients. The use of correlation is misleading.” The 1986 Bland–Altman article is highly cited, with over 42,000 citations in Google Scholar. Extensive descriptions of the methods are available in a number of subsequent articles that have described the Bland–Altman approach further [2–9].

Bland and Altman suggested plotting the differences between two methods against the mean of the two values and then using those data to derive the limits of agreement (LOA) between which 95% of the differences would be expected to fall. LOA are calculated as the mean difference \pm 2 standard deviations (SD). The well-known Bland–Altman plot shows the differences plotted against the mean of the two methods and displays the mean difference and the LOA. However, the method is more than just the plot. As explained succinctly by Altman and Bland in an “author reply” [10] to a 2002 letter: “The graph, which many think is the whole of our method, was intended as a visual check that the approach was reasonable and that the data were ‘well-behaved.’ Thus the graph shows whether the variability of differences between methods is roughly constant across the range of measurement, but the key element of the approach is to examine and summarize the individual differences between the two methods. Indeed, in our original paper we included histograms of these differences. This distribution should be approximately normal, and (apart from occasional outliers) this is usually what we see.” Bland and Altman’s method shows how to derive the LOA but did not propose an exact value for the acceptable degree of agreement, which they felt should be derived in advance if possible and based on clinical considerations for each specific application.

A number of studies [10–15] have evaluated the reporting of Bland–Altman analyses in method comparisons from anesthesiology, medical instrumentation, and clinical chemistry. These studies have identified limitations in how such analyses are reported. Commonly

reported deficiencies include lack of confidence intervals for the LOA and lack of prespecified criteria for acceptable LOA. For example, Mantha et al. [13] reviewed 42 articles in anesthesiology journals that had used Bland–Altman analysis for method comparisons and found that although 38 included correct Bland–Altman plots, only three specified acceptable LOA in advance and only one provided confidence intervals for the LOA. Dewitte et al. [10] reviewed 95 articles in the journal *Clinical Chemistry* and found that only 67 presented the LOA and only two compared the LOA with prespecified acceptable levels; they did not assess the issue of confidence intervals.

A number of systematic reviews have compared self-reported and measured weight and height or body mass index (BMI) calculated from self-reported or measured weight and height data [16–24]. These reviews, however, have not addressed the use of Bland–Altman analyses to assess agreement. Our objective is to evaluate the reporting of Bland and Altman methods in studies of self-reported versus measured weight and height or of BMI values calculated from those data.

Methods

Identification of articles

We used a variety of approaches to identify articles that compared self-reported and measured anthropometric values and used Bland–Altman methods. We included only studies in English that studied adults or adolescents, rather than children, and excluded studies that were published before 1986 (the date of the Bland–Altman publication in the *Lancet*). We examined the reference lists of systematic reviews that compare self-reported and measured weight and height or derived values such as BMI calculated from self-reported weight and height [16–24]. We repeated through May 2019 the PubMed search strategies described in two of these systematic reviews [21, 22]. We examined reference lists of papers cited in prior reviews and reference lists of epidemiologic studies of BMI and health outcomes that used self-reported heights and weights. We searched “related articles” in PubMed. We searched PubMed for articles that had cited Bland–Altman articles on method comparisons and then searched those results. We found 394 articles that compared self-reported with measured weight or height data or BMI calculated from self-reported data with BMI calculated from measured data. The complete list of the 394 studies is shown in Supplemental Table S1. Of those, we identified 72 published articles that used some form of Bland–Altman analysis.

Selection of criteria

We scored the 72 studies that used Bland–Altman analyses on the following six reporting criteria: assessment of the normality of the distribution of differences, a complete and correctly labeled Bland–Altman plot displaying the mean difference and LOA, numerical values and confidence intervals, standard errors, or SD for mean difference, numerical values of LOA, confidence intervals for LOA, and prespecified criteria for acceptable LOA. These criteria were derived as a modified version of the criteria that were suggested by Abu-Arafah et al. [11] in the context of anesthesiology research. Abu-Arafah et al. noted that the recommendations for adequate reporting in previous reviews of studies using the Bland–Altman approach were often not provided as an explicit list of specific requirements. Abu-

Arafah et al. performed a systematic search of the literature and compiled a list of 13 key features, based on the recommendations from 64 articles. Our criteria are a slightly modified version of that list. We did not include the criteria related to statistical software or the criteria related to repeated measures, because repeated measures have little applicability for studies of self-reported weight and height.

These six criteria provide information that can be used to evaluate the degree of agreement between two methods. The mean difference provides an estimate of bias. Assessment of normality may indicate the need for a transformation of the data before calculating the LOA. If possible, the acceptable limit of agreement should be prespecified. The LOA can be compared against the prespecified criteria to evaluate whether the degree of agreement is acceptable. Confidence intervals indicate the statistical uncertainty of the estimates and their generalizability to other populations.

All criteria were assessed by one author (KF), and the coauthors BG and JI each independently assessed a random sample of 20 articles. Consensus was reached on all scores. Furthermore, as we perused the eligible articles, we noted whether there were any other notable patterns of misuse and of interpretation for the presented Bland–Altman analyses.

Results

Eligible studies

Among the 394 screened studies, only 72 (18%) used some form of Bland–Altman analysis. A list of these 72 studies is shown in Supplemental Table 2. Of these 72 studies, 4 were published before 2000, 12 were published from 2000 to 2009, 34 were published from 2010 through 2014, and 22 were published from 2015 through May 2019. The 72 studies were highly heterogeneous, including studies from Europe [25], US/Canada [13], Australia/NZ [9], Brazil [7], Japan [4], and China [3], and one each from Mexico, Peru, and Malaysia. We did not attempt to evaluate the overall quality of the studies or summarize the results. There were numerous methodological differences among studies having to do with study design and timing of reports and measurements. A number of different approaches were used. For example, in several studies [26, 27], a subsample for measurement was selected from a larger group that had already provided self-reported weights and heights. In some studies [28, 29] participants were given specific instructions about measuring their own height and weight at home before providing reported values, which were then checked against a clinical examination. In another study [30], participants were asked to report weight and height and to indicate at the same time whether they wanted to have their weight and height subsequently measured.

Groups under study included occupational groups (bank employees, firefighters, public servants, nurses, and physicians), patients (bariatric surgery patients, HIV/AIDS patients, preoperative patients, emergency department patients, and breast cancer patients), participants in intervention studies (e.g., a randomized trial examining the effectiveness of two lifestyle-based weight loss interventions), and general population samples. The goals of these studies were also varied, ranging from general issues to highly specific applications.

For example, one study [31] wanted to evaluate the validity of self-reported data obtained from mailed surveys to derive BMI estimates, another study [32] sought to evaluate the use of self-reported data for Canadian health surveys, another [25] assessed the effects of using self-reported weight and height in an algorithm to detect left ventricular hypertrophy, and a fourth [33] evaluated the use of self-reported data to detect malnutrition in preoperative patients.

Criteria for use of Bland–Altman analyses

The distribution of overall scores is shown in Fig. 1. On the basis of six reporting criteria, no study satisfied more than four criteria. The number and percent of studies meeting each criterion is shown in Table 1. The overall scores and single criteria for each study are shown in Supplemental Table S2.

Of the 17 articles without complete Bland–Altman plots, three articles had no plot, and in eight articles, the plot was simply a plot of differences plotted against means and not the plot described by Bland and Altman. The remainder had additional lines that were not clearly identified.

Of the articles with prespecified criteria, two [34, 35] used criteria that compared the LOA with the SD of the measured values. Zhou et al. [35] compared the upper and lower LOA with the SD of the measured values, concluding that if the LOA were wider than 1 SD of the measured values (as was the case for height and weight) the agreement was “fair” and if wider than 2 SD of measured values (as was the case for BMI), the agreement was “poor.” Yoshitake et al. [34] similarly compared the LOA with the measured SD, concluding that for height, weight, and BMI, the agreement was “good” because the LOA were lower than 1 SD and sufficiently small to be considered acceptable. These articles did not provide further references or explanations for this practice. The third article by Villarini et al. [29] specified a difference of 10% or less as acceptable, citing an article by Barrios et al. [36] as justification.

Additional issues noted with use and interpretation of Bland–Altman analyses

We observed a number of issues while reviewing these articles. One problem is a circular interpretation of the LOA. Bland and Altman recommended that researchers decide, preferably in advance, on what constituted an acceptable degree of agreement. The LOA are chosen by design to cover 95% of the differences; thus, as long as the differences are approximately normally distributed or appropriately transformed, it would be expected that most of the differences fall within the LOA. A number of articles in our sample used circular reasoning to suggest that if most of the differences fell within the LOA, this indicated good agreement [37, 38]. For example, Pursey [38] stated: “at the group level, the majority of values fell within the LOA (2 SD) indicating a fairly good level of agreement.” Similarly, according to Krakowiak et al.: “the two sets of measurements were considered to be in good agreement if 95% of the differences were within 2 SD of the mean difference (i.e., the limits of agreement)” [37].

An issue that arose in a number of articles in our sample had to do with the distinction between the use of self-reported data for population prevalence estimates or the use for

individual assessments. Although precise criteria were not specified, a number of the studies in our sample concluded that the LOA indicated poor agreement at the individual level and thus that BMI may be valid for population estimates but not for individual assessments. Powell-Young [39] concluded: “the use of self-reported data for determining dosing parameters, identifying authentic association, tracking longitudinal weight changes, measuring obesity related morbidity, or clinical trials surveillance is not recommended.” Fonseca [40] opined that “at an individual level, self-reported BMI is not a valid screening tool, specifically in young adolescents.” Kee [41] felt that the “95% LOA ... indicated substantial discrepancies between self-reported and direct measurements method at the individual level” and concluded that self-reported weight and height were “not for assessing nutritional status at the individual level.” Olivarius et al. [42] concluded that “the high variability of the differences between recalled and measured weights makes it impossible to draw inferences on an individual level.” Zhou et al. [35] opined that “reported weight and height does not have an acceptable agreement with measured data. Therefore we do not recommend the application of self-reported weight and height to screen for overweight adolescents in China. Reported data could be considered for use in surveillance systems and epidemiology studies with caution. Any use of self-reported height and weight data from adolescents in future research studies should be justified with supporting pilot data validating such measures.” De Vriendt [43] felt that the results indicated “limited usefulness on an individual level.” Yoong et al. [44] found “substantial discrepancies at the individual level” but felt self-reported data were adequate for population trends. According to Pasalich et al. [45] “use of self-reported anthropometric data may be used more appropriately for describing overall distribution in population studies than for monitoring changes at an individual level.” In contrast, McAdams et al. [46] concluded that “the accuracy of self-reported BMI is sufficient for epidemiological studies using disease biomarkers, although inappropriate for precise measures of obesity prevalence.”

Discussion

The purpose of the Bland–Altman article was to discuss methods of comparing two measures of the same quantity, with the goal of assessing the agreement between the two methods. It appears, however, that the Bland–Altman approach is used only in a minority of studies of self-reported height or weight compared with measured height and weight. In our sample of 394 published articles that compared self-reported and measured height, weight, or BMI, 72 (18%) used some form of Bland–Altman methods. Even allowing for limitations in our sample, which may have omitted some relevant articles, this suggests that a relatively small proportion use such methods.

Even in the studies that nominally used Bland–Altman methods, there were deficiencies. We observed many of the same problems identified by others in other applications of this method in different fields such as clinical chemistry and anesthesiology [10–15]. These problems included incomplete plots, lack of calculation of LOAs, misunderstanding of LOAs, and lack of criteria to evaluate LOAs. Berthelsen and Nilsson [12] found that 14% of 50 papers in *Acta Anaesthesiologica Scandinavica* reported confidence limits for the LOA and 4% used prespecified criteria. Chhapola et al. [14] found that only 6% of their sample of 50 articles on comparisons of laboratory measurements reported confidence limits but, in

contrast to other studies, found that 74% used prespecified criteria. This difference may reflect the field of clinical chemistry with better-defined acceptable limits for laboratory analytes. They concluded that “despite its simplicity, [Bland–Altman analysis] appears not to be completely understood by researchers, reviewers and editors of journals.”

With the exception of the three studies mentioned above, the 72 studies that reported using Bland–Altman methods did not prespecify clinical criteria for acceptable limits of LOA. According to Bland and Altman [1]: “how far apart measurements can be without causing difficulties will be a question of judgment. Ideally, it should be defined in advance to help in the interpretation of the method comparison and to choose the sample size.” In the absence of prespecified or defined criteria, similar data can lead to different conclusions. A comparison is shown in Table 2 for self-reported versus measured weight for six roughly comparable studies of women in the UK and the United States. One of these studies calculated LOA and the other five contained sufficient information that we could calculate the LOA. All show similar high correlations of 0.97 or 0.98. Mean differences were similar, ranging from 0.9 to 1.8 kg. LOA were also similar, covering a range of roughly 10 kg, with the smallest range being from –3.4 to 6.2 kg and the largest being from –6.4 to 8.2 kg; however, the conclusions varied. One study with LOA for weight of –4 to 6 kg concluded that self-reported weight data were not valid for use in epidemiological studies requiring accuracy at the level of the individual. Another concluded that caution should be used when applying these results to individuals. A study with LOA for weight from –3.4 to 6.2 kg concluded that self-reported weight had been shown to be valid for studying associations. The other studies also concluded that self-reported weight was “suitable” or “relatively valid” or “reasonably valid.” However, there is no clear basis for any of these conclusions and no criteria were specified. Berthelsen and Nilsson [12] described the lack of prespecified criteria as “like betting after the race” and noted that with this approach it would be “straightforward for a researcher to “prove” his or her personal prejudice.”

Some studies that did not use Bland–Altman methods have mentioned defined limits. Steventon et al. [47] felt that for telemonitoring of heart failure patients, a change of even 2 pounds could be considered important [47]. Bowring et al. [48] defined accurate report of height and weight as <2 cm and <2 kg difference between self-report and measured values, based in part on results from Brestoff et al. [49].

Our evaluation has some potential limitations. Some studies that used and reported Bland–Altman methods, including studies published in languages other than English, may have been missed. There may be some studies for which Bland–Altman methods would not have been applicable given the goals of the study. We did not attempt to evaluate this. Our sample of studies that compared self-reported with measured weight and height was not derived from a de novo systematic review but instead used previous systematic reviews and their updates, along with other sources, to identify relevant studies.

Conclusions

Bland–Altman methods appear to be infrequently used in studies comparing measured with self-reported weight, height, or BMI, and key information is missing in many of those that

use Bland–Altman methods. Completeness and standardization of reporting would be helpful and would enhance transparency. Future directions would be defining acceptable LOA values for different purposes and improving the reporting and application of Bland–Altman methods in studies of self-reported anthropometry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–10. [PubMed: 2868172]
2. Altman DG, Bland JM. Assessing agreement between methods of measurement. *Clin Chem*. 2017;63:1653–4. [PubMed: 28588121]
3. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003;22:85–93. [PubMed: 12858311]
4. Giavarina D. Understanding Bland Altman analysis. *Biochem Med*. 2015;25:141–51.
5. Hamilton C, Stamey J. Using Bland-Altman to assess agreement between two medical devices-don't forget the confidence intervals! *J Clin Monit Comput*. 2007;21:331–3. [PubMed: 17909978]
6. Ludbrook J. Confidence in Altman-Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol*. 2010;37:143–9. [PubMed: 19719745]
7. Phatak AG, Nimbalkar SM. Method comparison (agreement) studies: myths and rationale. *J Clin Diagn Res*. 2017;11:Ji01–3.
8. Twomey PJ. How to use difference plots in quantitative method comparison studies. *Ann Clin Biochem*. 2006;43:124–9. [PubMed: 16536914]
9. Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. *Theriogenology*. 2010;73:1167–79. [PubMed: 20138353]
10. Dewitte K, Fierens C, Stockl D, Thienpont LM. Application of the Bland-Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. *Clin Chem*. 2002;48:799–801. Author reply 801–792. [PubMed: 11978620]
11. Abu-Arafeh A, Jordan H, Drummond G. Reporting of method comparison studies: a review of advice, an assessment of current practice, and specific suggestions for future reports. *Br J Anaesth*. 2016;117:569–75. [PubMed: 27799171]
12. Berthelsen PG, Nilsson LB. Researcher bias and generalization of results in bias and limits of agreement analyses: a commentary based on the review of 50 *Acta Anaesthesiologica Scandinavica* papers using the Altman-Bland approach. *Acta Anaesthesiol Scand*. 2006;50:1111–3. [PubMed: 16987340]
13. Mantha S, Roizen MF, Fleisher LA, Thisted R, Foss J. Comparing methods of clinical measurement: reporting standards for bland and altman analysis. *Anesth Analg*. 2000;90:593–602. [PubMed: 10702443]
14. Chhapola V, Kanwal SK, Brar R. Reporting standards for Bland-Altman agreement analysis in laboratory research: a cross-sectional survey of current practice. *Ann Clin Biochem*. 2015;52:382–6. [PubMed: 25214637]
15. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. *PLoS ONE*. 2012;7:e37908. [PubMed: 22662248]
16. Bowman RL, DeLucia JL. Accuracy of self-reported weight: a meta-analysis. *Behav Ther*. 1992;23:637–55.
17. Connor Gorber S, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev*. 2007;8: 307–26. [PubMed: 17578381]

18. 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;48:338–45. [PubMed: 14526347]
19. He J, Cai Z, Fan X. Accuracy of using self-reported data to screen children and adolescents for overweight and obesity status: a diagnostic meta-analysis. *Obes Res Clin Pract*. 2017;11:257–67. [PubMed: 28389205]
20. Headen I, Cohen AK, Mujahid M, Abrams B. The accuracy of self-reported pregnancy-related weight: a systematic review. *Obes Rev*. 2017;18:350–69. [PubMed: 28170169]
21. Maukonen M, Mannisto S, Tolonen H. A comparison of measured versus self-reported anthropometrics for assessing obesity in adults: a literature review. *Scand J Public Health*. 2018;46:10.1177/1403494818761971.
22. Seijo M, Minckas N, Cormick G, Comande D, Ciapponi A, Belizan JM. Comparison of self-reported and directly measured weight and height among women of reproductive age: a systematic review and meta-analysis. *Acta Obstet Gynecol Scand*. 2018;97:429–39. [PubMed: 29430640]
23. Sherry B, Jefferds ME, Grummer-Strawn LM. Accuracy of adolescent self-report of height and weight in assessing overweight status: a literature review. *Arch Pediatr Adolesc Med*. 2007;161:1154–61. [PubMed: 18056560]
24. De Rubeis V, Bayat S, Griffith LE, Smith BT, Anderson LN. Validity of self-reported recall of anthropometric measures in early life: a systematic review and meta-analysis. *Obes Rev*. 2019;20:1426–40. [PubMed: 31184422]
25. Cuspidi C, Negri F, Giudici V, Muiasan ML, Grandi AM, Ganau A, et al. Self-reported weight and height: implications for left ventricular hypertrophy detection. An Italian multi-center study. *Clin Exp Hypertens*. 2011;33:192–201. [PubMed: 21446894]
26. Wang Z, Patterson CM, Hills AP. A comparison of self-reported and measured height, weight and BMI in Australian adolescents. *Aust N Z J Public Health*. 2002;26:473–8. [PubMed: 12413294]
27. Bes-Rastrollo M, Sabate J, Jaceldo-Siegl K, Fraser GE. Validation of self-reported anthropometrics in the Adventist Health Study 2. *BMC Public Health*. 2011;11:213. [PubMed: 21466678]
28. Celis-Morales C, Livingstone KM, Woolhead C, Forster H, O'Donovan CB, Macready AL, et al. How reliable is internet-based self-reported identity, socio-demographic and obesity measures in European adults? *Genes Nutrition*. 2015;10:28. [PubMed: 26143178]
29. Villarini M, Acito M, Gianfredi V, Berrino F, Gargano G, Somaini M, et al. Validation of self-reported anthropometric measures and body mass index in a subcohort of the dianaweb population study. *Clin Breast Cancer*. 2019;19:e511–8. [PubMed: 31182401]
30. Xie YJ, Ho SC, Liu ZM, Hui SS. Comparisons of measured and self-reported anthropometric variables and blood pressure in a sample of Hong Kong female nurses. *PLoS ONE*. 2014;9:e107233. [PubMed: 25222707]
31. Burton NW, Brown W, Dobson A. Accuracy of body mass index estimated from self-reported height and weight in mid-aged Australian women. *Aust N Z J Public Health*. 2010;34:620–3. [PubMed: 21134066]
32. Elgar FJ, Stewart JM. Validity of self-report screening for overweight and obesity. Evidence from the Canadian Community Health Survey. *Can J Public Health*. 2008;99:423–7. [PubMed: 19009930]
33. Haverkort EB, de Haan RJ, Binnekade JM, van Bokhorst-de van der Schueren MA. Self-reporting of height and weight: valid and reliable identification of malnutrition in preoperative patients. *Am J Surg*. 2012;203:700–7. [PubMed: 22365153]
34. Yoshitake N, Okuda M, Sasaki S, Kunitsugu I, Hobara T. Validity of self-reported body mass index of Japanese children and adolescents. *Pediatr Int*. 2012;54:397–401. [PubMed: 22168382]
35. Zhou X, Dibley MJ, Cheng Y, Ouyang X, Yan H. Validity of self-reported weight, height and resultant body mass index in Chinese adolescents and factors associated with errors in self-reports. *BMC Public Health*. 2010;10:190. [PubMed: 20384994]
36. Barrios P, Martin-Biggers J, Quick V, Byrd-Bredbenner C. Reliability and criterion validity of self-measured waist, hip, and neck circumferences. *BMC Med Res Methodol*. 2016; 16:49. [PubMed: 27145829]

37. Krakowiak P, Walker CK, Tancredi DJ, Hertz-Picciotto I. Maternal recall versus medical records of metabolic conditions from the prenatal period: a validation study. *Matern Child Health J*. 2015;19:1925–35. [PubMed: 25656730]
38. Pursey K, Burrows TL, Stanwell P, Collins CE. How accurate is web-based self-reported height, weight, and body mass index in young adults? *J Med Internet Res*. 2014;16:e4. [PubMed: 24398335]
39. Powell-Young YM. The validity of self-report weight and height as a surrogate method for direct measurement. *Appl Nurs Res*. 2012;25:25–30. [PubMed: 20974100]
40. Fonseca H, Silva AM, Matos MG, Esteves I, Costa P, Guerra A, et al. Validity of BMI based on self-reported weight and height in adolescents. *Acta Paediatr*. 2010;99:83–88. [PubMed: 19878130]
41. Kee CC, Lim KH, Sumarni MG, Teh CH, Chan YY, Nur Hafizah MI, et al. Validity of self-reported weight and height: a cross-sectional study among Malaysian adolescents. *BMC Med Res Methodol*. 2017;17:85. [PubMed: 28577547]
42. Olivarius NF, Andreasen AH, Loken J. Accuracy of 1-, 5- and 10-year body weight recall given in a standard questionnaire. *Int J Obes Relat Metab Disord*. 1997;21:67–71. [PubMed: 9023604]
43. De Vriendt T, Huybrechts I, Ottevaere C, Van Trimpont I, De Henauw S. Validity of self-reported weight and height of adolescents, its impact on classification into BMI-categories and the association with weighing behaviour. *Int J Environ Res Public Health*. 2009;6:2696–711. [PubMed: 20054463]
44. Yoong SL, Carey ML, D’Este C, Sanson-Fisher RW. Agreement between self-reported and measured weight and height collected in general practice patients: a prospective study. *BMC Med Res Methodol*. 2013;13:38. [PubMed: 23510189]
45. Pasalich M, Lee AH, Burke L, Jancey J, Howat P. Accuracy of self-reported anthropometric measures in older Australian adults. *Australas J Ageing*. 2014;33:E27–32. [PubMed: 24520988]
46. McAdams MA, van Dam RM, Hu FB. Comparison of self-reported and measured BMI as correlates of disease markers in US adults. *Obesity*. 2007;15:188–96. [PubMed: 17228047]
47. Steventon A, Chaudhry SI, Lin Z, Mattered JA, Krumholz HM. Assessing the reliability of self-reported weight for the management of heart failure: application of fraud detection methods to a randomised trial of telemonitoring. *BMC Med Inform Decis Mak*. 2017;17:43. [PubMed: 28420352]
48. Bowring AL, Peeters A, Freak-Poli R, Lim MS, Gouillou M, Hellard M. Measuring the accuracy of self-reported height and weight in a community-based sample of young people. *BMC Med Res Methodol*. 2012;12:175. [PubMed: 23170838]
49. Brestoff JR, Perry IJ, Van den Broeck J. Challenging the role of social norms regarding body weight as an explanation for weight, height, and BMI misreporting biases: development and application of a new approach to examining misreporting and misclassification bias in surveys. *BMC Public Health*. 2011;11:331. [PubMed: 21592341]
50. 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;31:169–74. [PubMed: 12006304]
51. Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr*. 2002;5:561–5. [PubMed: 12186665]
52. Wright FL, Green J, Reeves G, Beral V, Cairns BJ. Validity over time of self-reported anthropometric variables during follow-up of a large cohort of UK women. *BMC Med Res Methodol*. 2015;15:81. [PubMed: 26450616]
53. Pirie P, Jacobs D, Jeffery R, Hannan P. Distortion in self-reported height and weight data. *J Am Diet Assoc*. 1981;78:601–6. [PubMed: 7252014]
54. Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC. Validity of self-reported waist and hip circumferences in men and women. *Epidemiology*. 1990;1:466–73. [PubMed: 2090285]
55. Luo J, Thomson CA, Hendryx M, Tinker LF, Manson JE, Li Y, et al. Accuracy of self-reported weight in the Women’s Health Initiative. *Public Health Nutr*. 2019;22:1–10. [PubMed: 30595139]

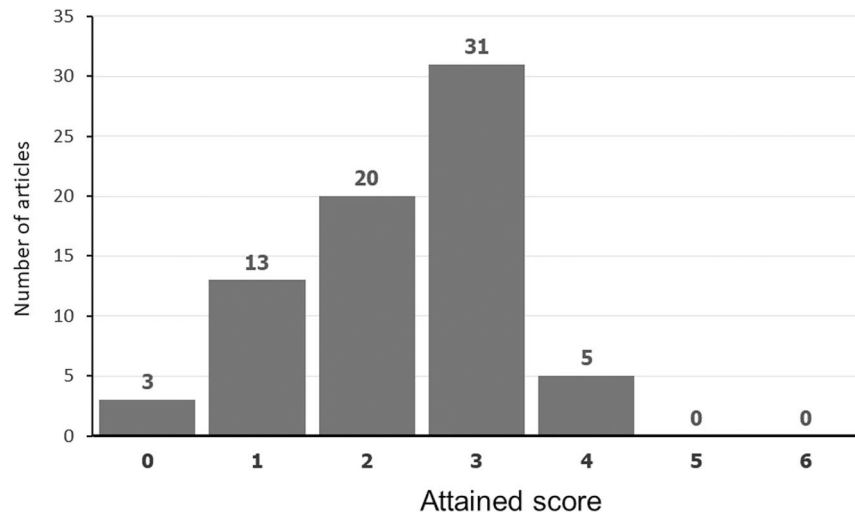


Fig. 1.
Distribution of scores

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Of 72 studies using Bland–Altman methods, number and percent meeting specified criteria

Criteria	Studies meeting criterion	
	Number	Percent
Assessment of the normality of the distribution of differences	4/72	6
A complete and correctly labeled Bland–Altman plot displaying the mean difference and limits of agreement (LOA)	55/72	76
Numerical values and confidence intervals, standard errors or standard deviations for mean difference	64/72	89
Numerical values of LOA	37/72	51
Confidence intervals for LOA	3/72	4
Prespecified criteria for acceptable LOA	3/72	4

Table 2

Comparison of conclusions for self-reported versus measured weight from six studies of women in the UK and the United States with similar values of correlations, mean differences, and limits of agreement (LOA)

Author (reference)	Study sample	Correlation	Mean difference (kg)	LOA (kg)		Conclusion
				Lower	Upper	
Lawlor et al. [50]	1310 women from the British Women's Heart & Health Study, recruited 1999–2000 at ages 60–79 years	0.98	1.0	-4	6.0	Self-report of weight should not be relied upon in prospective epidemiological studies or clinical practice when accuracy at the level of the individual is required.
Spencer et al. [51] ^a	2938 women from the Oxford-European Prospective Investigation into Cancer (EPIC) study in England recruited in 1993–1999 at ages 35–76 (median 52)	>0.9	1.4	-3.4	6.2	Self-reported ... weight data have been shown here to be valid for identifying associations in epidemiological studies. In analyses where anthropometric factors are the primary variables of interest, measurements in a random sample of the study population can be used to improve the accuracy of estimates of height, weight, and BMI.
Wright et al. [52] ^a	3999 women from the Million Women Study in England and Scotland, recruited between 1996 and 2001 at ages 50–64, measured in 2008	0.97	1.1	-5.0	7.2	Overall, however, we found that self-reported weight ... [is] suitable for use in epidemiological analyses with long-term follow-up.
Pirie et al. [53] ^{a,b}	1059 women ages 40–59 from the Minnesota sample of the Lipid Research Clinics Prevalence Program ages 40–59, measured ca. 1980	0.97	1.8	-3.8	7.4	Caution should be used when applying these results to individuals. The standard deviations of reporting discrepancy are high relative to the mean discrepancy ... reported height and weight are not accurate as measures of true height and weight.
Rimm et al. [54] ^{a,b}	140 women from the US Nurses' Health Study, ages 41–64 at measurement in 1987	0.97	1.1	-4.4	6.5	Self-reported ... weight measurements appear reasonably valid.
Luo et al. [55] ^{a,b}	75,336 women from the US Women's Health Initiative Observational Study ages 51–79 at baseline in 1993–1998	0.97	0.9	-6.4	8.2	[T]his large prospective study confirmed previously reported results that women demonstrate relatively valid estimates of body weight.

^aLOA calculated as mean difference \pm 2 SD based on published mean difference, sample size, and SE or SD

^bConverted from lb. to kg