Comparison of self-reported and directly measured weight and height among
 women of reproductive age: a systematic review and meta-analysis

- 3 Running Title: Self-report weight and height in women
- 4 Mariana Seijo^{1§*}, Nicole Minckas^{1§}, Gabriela Cormick^{1,2}, Daniel Comandé³, Agustín
- 5 Ciapponi³, José M. Belizan¹.
- 6
- ⁷ ¹ Department of Mother and Child Health Research, Institute for Clinical Effectiveness
- 8 and Health Policy (IECS-CONICET), Buenos Aires, Argentina.
- ² Department of Human Biology, Faculty of Health Sciences, University of Cape Town.
- ³ Argentine Cochrane Center, Institute for Clinical Effectiveness and Health Policy
- 11 (IECS-CONICET), Buenos Aires, Argentina.
- 12 [§]: Both authors contributed equally to this work
- 13 *<u>Corresponding author</u>:
- 14 Department of Mother and Child Health Research, Institute for Clinical Effectiveness
- 15 and Health Policy (IECS-CONICET), Buenos Aires, Argentina.
- 16 Dr. Emilio Ravignani 2024 (C1414CPV), Buenos Aires, Argentina
- 17 (+54-11) 4966-0082
- 18 Email: <u>mseijo@iecs.org.ar</u>

ce

- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26

27	
28	
29	Conflict of Interest Statements: No conflict of interest was declared.
30	
31	
32	
33	
34	
35	5
36	
37	
38	
39	
40	×C
41	
42	
43	
44	RCCR
45	
46	
47	
48	
49	

50 Full Abstract

Introduction: The use of self-report as a strategy for collecting data of women's weight 51 and height is broadly spread both in clinical practice and epidemiological studies. This 52 study aimed to compare self-reported and directly measured weight and height among 53 women of reproductive age. Material and methods: In July 2015 we searched 54 MEDLINE, EMBASE, COCHRANE, CINHAL, LILACS and grey literature. We 55 included women of reproductive age (12 to 49 years old) independently of their weight 56 or height at the time of the study. Women with any condition that implies regular track 57 of their weight (e.g., eating disorder) were excluded. Two reviewers independently 58 selected, extracted and assessed the risk of bias of the studies. We used RevMan 5.3 to 59 perform the meta-analysis. Heterogeneity was assessed using the I^2 statistic. Results: 60 Following eligibility assessment, 21 studies including 18 749 women met the inclusion 61 criteria. The results of the meta-analysis showed an underestimation of weight by -62 0.94kg (95%CI -1.17, -0.71kg; p<0.0001; I²=0%) in the overall sample and an 63 overestimation of height by 0.36cm (95%CI 0.20, 0.51; p<0.0001; $I^2=35\%$) based on 64 self-reported as compared to directly measured values. Conclusions: This review shows 65 that self-reported weight and height of women of reproductive age is slightly different 66 than direct measures. We consider that the magnitude at which self-reported data over 67 or underestimates the real value is negligible regarding clinical and research use. 68

Key words: Self-Assessment, Body Weights and Measures, Body Weight, BodyHeight, Body Mass Index, Women, Reproductive Age.

71

72 List of abbreviation:

- 73 BMI: Body mass index
- 74 CI: Confidence Interval
- 75 DM: Direct measured
- 76 SD: Standard Deviation
- 77 SE: Standard Error
- 78 SR: Self-reported
- 79 WHO: World Health Organization

Key messages: Self-reported weight and height in women of reproductive age is a
measure that closely estimates the real values and can be used as proxy both in clinical
and research evaluations related to reproductive health.

Accepted manuscrit

83 Introduction

Body mass index (BMI) is a simple and useful indicator to classify individuals as 84 healthy or at risk according to their weight and height ⁽¹⁾. Traditional anthropometric 85 measures such as weight and BMI are often used in epidemiological studies to assess 86 changes in population health and nutritional status ⁽²⁾. Regarding women's health, BMI 87 prior to pregnancy requires strict attention as it can be a risk factor not only for women, 88 but also for future generations⁽³⁾. Because of this, the International Federation of 89 Gynecology and Obstetrics (FIGO) emphasizes the need to control pre-conceptional 90 body weight and BMI to prevent abnormal values that can impact significantly on 91 maternal and neonatal health outcomes $^{(3)}$. 92

Anthropometric measures are often gathered through self-administered questionnaires. 93 This data collection method has the advantages of being quick, easy to administer, and 94 cost-effective when working with large samples, or when individuals are spread over 95 large areas ⁽⁴⁾. In research, the self-report of height or weight is highly used in 96 descriptive studies to save significant amount of time and resources ⁽⁵⁻⁸⁾. In clinical 97 practice, self-reported measures of weight are also a useful strategy to determine 98 historical weights; for example, self-report allows for estimation of pregnancy weight 99 gain that would otherwise be difficult due to the variable stages in which the first 100 antenatal visit occurs. Despite these advantages, the utility of self-reported measures has 101 been questioned, particularly when it relates to anthropometric measures. There is a 102 global preconceived idea that participants tend to overestimate their height and 103 underestimate their weight, resulting in a lower estimate of BMI⁽⁴⁾. The greatest hazard 104 of unreliable reporting of weight and height is the inaccurate estimation of the 105 106 prevalence of overweight and obesity, which can result in unsupported decision-making (4). 107

108 It is vital to have an up-to-date systematic review on this topic in order to reduce the 109 risk of bias when reporting the results of a study. Any important difference between 110 self-reported and directly measured data found should be taken into consideration when 111 selecting data collection methods for future studies or clinical actions.

112 The objective of this review is to compare self-reported with directly measured weight 113 and height among women of childbearing age. The purpose of these meta-analyses is to give a summary estimate of the possible bias that can occur when using self-report as a data collection method.

116 Methods

117 We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses

118 Statement (PRISMA Statement)^(9, 10) and the Meta-analysis Of Observational Studies in

119 Epidemiology (MOOSE) statement.

120 Criteria for considering studies for this review

We selected cross-sectional and prospective or historical cohort studies that compared individual self-reported with directly measured weight and height data. We included published or unpublished studies from 2000 onward that reported at least 20-paired values of self-reported and directly measured weight or height, or data of the difference between them. No language restriction was used.

We included healthy non-pregnant women of reproductive age, independent of their weight or height. We considered reproductive age to be from 12 to 49 years old. All methods to obtain a self-reported or directly measured weight and height were accepted. We excluded women with a disease or condition that implied regular monitoring or records of their weight, such as women following dietary plans or women with eating disorders.

Studies were included only if they expressed the outcome as "mean self-reported weight
or height", "mean directly measured weight or height" or "mean difference between
self-reported and directly measured weight or height".

135 Search methods for identification of studies

136 *Electronic searches*

A literature search for articles published from January 1st, 2000 to July 14th, 2015 was
conducted within the main international and regional databases, through generic and
academic internet searches, and through meta-search engines.

- 140 We searched records from the following databases:
- CENTRAL: The Cochrane Library (last available Issue 2015)
- MEDLINE (January 2000 to July 2015)

- EMBASE (January 2000 to July 2015)
- LILACS: Latin American and Caribbean Health Science Literature (January 2000 to July 2015).
- CINAHL (January 2000 to July 2015)

The simplified and complete search with filters in MEDLINE is described below; these
were adapted appropriately for each database (Supporting Information Appendix S1
Supplementary Methods). We also reviewed the reference lists of included studies for
potential additional studies.

- 151 *Data collection and analysis*
- 152 *Selection of studies*

All phases of the study selection and processing were completed using EROS® (Early 153 Review Organizing Software, IECS, Buenos Aires), a web-based platform designed for 154 the process of systematic reviews⁽¹¹⁾. As an initial screening, pairs of reviewers (MS, 155 NM) independently reviewed the articles, evaluating the titles and abstracts of identified 156 studies according to pre-specified criteria. Discrepancies were resolved by consensus of 157 the whole research team. Articles included after the initial evaluation were retrieved in 158 159 full text for a second screening to determine eligibility. Finally, the same reviewers independently extracted and assessed the risk of bias of each full text article. 160

161 *Data extraction and management*

We used a web-based spreadsheet to extract the information. One reviewer extracted data from the included studies and a second reviewer double-checked it to minimize potential errors. This process was piloted on 20 papers to refine it. Discrepancies were resolved by consensus of the whole team.

The information extracted from each study included author, publication year, type of 166 study, region and country of study, participant characteristics (age and education level), 167 sample size, methods to obtain directly measured weight and height (stadiometer, 168 169 anthropometer, or other type of measuring (tape or ruler, variety of scales), methods to 170 obtain self-reported weight and height (long distance survey, on-site interview, selfadministered questionnaire), time between collection of self-reported and directly 171 measured data, order of measures, ethical considerations, and outcomes (mean self-172 reported and directly measured weight or height or mean differences between self-173

reported and directly measured weight or height, and their standard deviation [SD]).
Authors of studies reporting incomplete information were contacted to provide missing
information. We waited for one month for the author's answer before excluding the
article.

178 Assessment of risk of bias and data analysis

The risk of bias of observational studies was assessed using a checklist of essential
items based on the STROBE (Strengthening the Reporting of Observational studies in
Epidemiology) ⁽¹²⁾. The STROBE essential checklist includes: selection of participants,
control of confounders, measurement of exposure and outcome, and conflict of interest.
Pairs of independent reviewers assessed the methodological quality. Discrepancies were
resolved by consensus of the whole team.

The null hypothesis when comparing self-reported and directly measured weight and 185 height stated no difference between methods (self-reported = directly measured). Those 186 measurements expressed in pounds or inches were transformed to kilograms and 187 centimeters, respectively, and the reported standard errors (SE) were converted to SD 188 using the following formula: $\sqrt{n} \times SE$. We performed a meta-analysis using the 189 continuous outcomes of all the studies that reported mean values of weight or height 190 using self-reported and directly measured methods. A summary estimate obtained from 191 192 the meta-analysis of a mean difference not equal to 0 would indicate that the use of selfreport affects positively or negatively on the measure compared to the use of direct 193 194 measurements; based on either difference, self-reported values could be defined as a weak method for data collection. We used RevMan 5.3⁽¹³⁾ to perform the meta-analysis 195 196 and to calculate the two-tailed P-values and 95% confidence intervals (CI).

We measured heterogeneity using the I² statistic as follows: low heterogeneity (I² less than 25%), moderate heterogeneity (I² between 25–75%), and high heterogeneity (I² greater than 75%).

For those studies that only reported mean differences between methods, we performed a generic inverse-variance meta-analysis, which considered mean difference and SE. To be able to include all the studies we used RevMan's calculator function to extract mean differences and SE for each of them. The resulting value indicated the directionality of the findings. A result under 1 indicated that the directly measured values were higher than the self-reported ones; a result above 1 indicated that the self-reported values were
higher than the direct measured ones; a value of 1 indicated no difference between
methods.

Pre-specified subgroup analyses by age, time between self-reported and direct measured
measurement (same day, different days), region of the study (Latin America &
Caribbean, Europe, North America, Oceania, Asia), self-report method (long distance
survey, self-administered questionnaire on-site, in-person interview) and women's BMI
were performed. For all the meta-analyses we used a random effect model to address
possible clinical or methodological heterogeneity between studies.

We compiled the age data into three groups: 1) 12 to 18 years, 2) 19 to 35 years and 3) 36 to 49 years. For studies in which age was grouped differently and data could not be disaggregated, we based our groups on the category to which the majority of study participants belonged. BMI was classified following WHO categories (underweight less than 18.5, normal weight 18.5 to less than 25, overweight 25 to less than 30 and obesity 30 or more)⁽¹⁾. The protocol was registered in PROSPERO, an international prospective register of systematic review protocols (Registration Number CRD42015029142).

221 **Results**

222 Description of studies

223 *Results of the search*

The search strategy retrieved 1638 references after removing duplicates. Of those, 1476 references were excluded by title and abstract, leaving 162. Two full texts were not found ^(14, 15) and 139 studies did not meet the inclusion criteria. After assessment, 21 studies with 18 749 women were included in the review (Figure 1). ⁽¹⁶⁻³⁶⁾

228 Included studies

Of the 21 included studies, six were from Latin America and the Caribbean (n=3470, 14.8% of the women), $^{(18, 29, 30, 32, 35, 36)}$ nine from Europe (n=8459, 36.2% of the women), $^{(16, 19-22, 24, 26, 28, 33)}$ and four from North America (n=8264, 35.3% of the women) $^{(17, 23, 25, 34)}$. We only included one article from Oceania $^{(31)}$ and one from Asia $^{(27)}$ (n=3206, 13.7% of the women). Regarding design, two of the included studies were prospective cohorts $^{(19, 33)}$ and the rest (n=19) were cross-sectional studies (Table 1) $^{(16-18, 20-32, 34-36)}$.

Eighteen studies reported details of the tools used for self-reported and directly 236 measured weight and height of participants (16-22, 24-29, 31, 33-36). For directly measured 237 data, height was most commonly measured by stadiometer, anthropometer, or some 238 type of measuring tape or ruler with an error between 0.1 to 0.5 cm, while weight was 239 measured by a variety of scales with an error of 0.1kg (balance beam, digital, or 240 portable). Twelve of the 21 studies used self-administered on-site questionnaires as the 241 self-reported method.^(17, 20-22, 24-27, 33-36) Three studies gathered information in an online 242 survey or via telephone (18, 19, 31), while three other studies performed an in-person 243 interview to obtain this data. ^(16, 28, 29) The remaining three studies did not report the type 244 of methods used ^(23, 30, 32). All the studies obtained the self-reported value prior to the 245 directly measured data (16-36). 246

From the included studies, nineteen reported mean value of self-reported and directly
measured weight and height ^(16-21, 23-28, 30-36). Two studies only reported mean difference
between methods, calculated as self-reported minus directly measured values ^(22, 29).
Only two studies showed data by women's BMI categories ^(28, 29).

251 Risk of bias Assessment

The risk of bias assessment found six studies with high risk of bias in the selection of participants $(29.0\%)^{(24, 25, 28, 30, 33, 36)}$ and two studies with the control of confounders (9.50%) (Supporting Information Table S1)^(30, 36).

255 Weight

According to the meta-analysis, we found that in the overall sample, the mean 256 difference between self-reported and direct measured data for women's weight was -257 258 ^{23-28, 30-36}. When analyzed by age subgroups, we found that self-reported weight was 259 lower than directly measured weight in women between 12 and 18 years (-1.05 260 [95%CI; -1.32, -0.78]; p<0.0001; I²=0%) and in women between 19 and 35 years (-1.04) 261 [95%CI; -1.86, -0.21]; p=0.001, I²=30%). However, in women from 36 to 49 years, 262 there was no statistically significant difference between methods (-0.26 [95%CI; -0.99, 263 0.44]; p=0.49; I²=0%) (Figure 2- Panel A). 264

The results by region were in the same direction in all three meta-analyzed regions (Figure 3 - Panel A). The difference between self-reported and directly measured weight was -1.14kg (95%CI; -1.67, -0.61; p<0.0001; $I^2=0\%$) in Latin America and the Caribbean, -1.02kg (95%CI; -1.68, -0.37]; p=0.002, $I^2=55\%$) in Europe; and -1.51kg (95%CI; -2.53, -0.48; p=0.004; $I^2=0\%$) in North America. We only found one study for Asia and one for Oceania, and they were not included in the meta-analysis ^(24, 28).

In the analysis by time of data collection we found that if obtained within the same day there was a -0.97kg (95%CI; -1.37, -0.57; P<0.001; I²=15%) difference between selfreported and directly measured weight. No statistically significant difference was found when collected on separate days (-1.64kg [95%CI; -4.30, 1.03]; p=0.23; I²=0%) (Figure 4 - Panel A).

We also evaluated the influence of the self-reported method used when compared to directly measured data (Supporting Information Figure S1 - Panel A). The analysis suggested that there was a negative difference if the information was gathered through a long-distance survey (-1.46kg [95%CI; -2.27, -0.64]; p=0.0004; I²=0%) or a selfadministered questionnaire on-site (-1.14kg [95%CI; -1.79, -0.48]; p=0.006; I²=54%). The difference was lower when gathered during an in-person interview (-0.27kg [95%CI; -0.80, 0.25]; p=0.74; I²=46%).

Only two studies classified their population according to BMI status of participants. We found that those who were overweight underestimated their weight by -0.39kg ([95%CI; -0.59, -0.19]; p=0.0001; $I^2=0\%$) ^(28, 29). We found no statistically significant results because of the high heterogeneity between the studies for the other three BMI categories (underweight, normal weight or obesity) (Figure 5 - Panel A).

As mentioned previously, two studies only reported mean difference between methods, 288 without specifying mean self-reported weight and mean direct measured weight ^(22, 29). 289 One study included women between 15 and 18 years ⁽²²⁾. The second study divided its 290 population into three subgroups: 20 to 24 years, 25 to 34 years, and 35 to 44 years⁽²⁹⁾. 291 We performed a separate analysis to evaluate if the results of these studies were 292 consistent with the directionality of the findings previously presented. We meta-293 analyzed these population subgroups and found an I^2 of 80% (Supporting Information 294 Figure S2 – Panel A). 295

296 *Height*

According to the meta-analysis, we found that in the overall sample, the mean 297 difference between self-reported and directly measured data for women's height was 298 0.36cm (18 studies; 13 744 participants; [95%CI; 0.20, 0.51]; p<0.0001; $I^2=35\%$) (16-21, 299 ^{23-28, 30-33, 35, 36)}. When analyzed by age, we found that self-reported height was higher 300 than directly measured height in all subgroups (Figure 2 - Panel B). In the subgroup of 301 age between 12 and 18 years the mean difference was 0.24cm ([95%CI; 0.04, 0.44]; 302 p=0.02; $I^2=54\%$); in the group between 19 and 35 years the mean difference was 303 0.57cm ([95%CI; 0.25, 0.89]; p<0.001; I²=0%); and in the subgroup women from 36 to 304 49 years the mean difference was 0.50cm ([95%CI; 0.09, 0.91]; p=0.02; I²=0%). 305

The analysis by region showed a significant mean difference between self-reported height and directly measured height of 0.63cm ([95%CI; 0.41, 0.85]; p<0.0001; I²=0%) in Europe. No statistical differences were found in the Americas (North America: -0.62cm [95%CI; -1.30, 0.06; p=0.08; I²=0%], or Latin America and the Caribbean: 0.43cm ([95%CI; -0.07, 0.92]; p=0.09; I²=11%) (Figure 3 - Panel B). Two studies were excluded from the meta-analysis because each was the only reference from their region (27, 31).

In the analysis by time of data collection we found that if obtained within the same day, the difference was 0.53cm (95%CI; 0.20, 0.85; p=0.001; I2=43%). No significant difference was found when obtained on separate days (0.60 cm [95%CI; -0.83, 2.04]; p=0.41; I2=0%) (Figure 4 - Panel B).

We also evaluated the influence of the specific self-reported method used when compared to directly measured height data (Supporting Information Figure S1- Panel B). The analysis showed a significant difference if the information was gathered through a long-distance survey (0.55cm [95%CI; 0.00, 1.09]; p=0.05; I²=0%) or in an in-person interview (0.65cm [95%CI; 0.28, 1.02]; p=0.0005; I²=38%). No statistically significant difference was found when the data was gathered through an on-site selfadministered questionnaire (0.10cm [95%CI; -0.68, 0.47]; p=0.72; I2=70%).

The high heterogeneity found between studies in the subgroup analysis based on women's BMI categories prevented us from obtaining an estimate difference between self-reported and directly measured height ^(28, 29) (Figure 5 - Panel B). The separate analysis for the two studies $^{(27, 31)}$ reporting only the mean difference between methods found that the results were consistent with the findings previously presented, and showing self-reported height higher than direct measurements. There was no heterogeneity between studies (I²=0%) (Supporting Information Figure S2 – Panel B).

332

333

334 Discussion

The results of this review showed an overall underestimation of weight (-0.94kg) and an overestimation of height (+0.36cm) when comparing self-reported to directly measured values in women of reproductive age.

In the pre-specified subgroup analyses, the findings remained consistent. We found that 338 339 women aged 12 to 35 years under-reported their weight by 0.78kg to 1.17kg. Older women also under-reported their weight, but the difference was not statistically 340 significant in this age group. The underestimation of self-reported weight was found 341 throughout all studied regions reaching a mean difference between self-reported and 342 direct measured weight as high as 1.50 kg in North America. Few studies presented data 343 in overweight women; the results on weight were similar to normal weight women. It 344 was not possible to estimate the differences by underweight or obese subgroups. 345

We found that the underestimation of weight persisted if data was collected through an on-site self-administered questionnaire or a long-distance survey (online or via telephone); however, when self-reported data was collected by on-site in-person interviews, this underestimation was lower and not statistically significant.

Regarding height, the results showed a consistent overestimation throughout all age groups. These findings were also observed in studies from Europe and North America, but not in those from Latin America and the Caribbean. The overestimation in height persisted when collected through an on-site in-person interview or long distance survey; however, there was no statistically significant difference with directly measured values when using an on-site self-administered questionnaire. Our results confirmed the data published by Gorber et al ⁽³⁷⁾ in the general population and updated by Engstrom et al ⁽³⁸⁾ results from 2002 to 2015. All these studies showed an underestimation of weight
and overestimation of height. In our study, as well as that of Gorber and Engstrom'
reviews, the standard deviations were large in all included studies, suggesting
significant variability between women in the accuracy of self-reported height and
weight measurements.

One of the authors carried out 3 pilot tests of search strategies MEDLINE to explore the potential sensitivity and specificity of the electronic searches. We assume that the risk of publication bias is low (Supporting Information Table S1). Poor reporting of studies was the major problem found when assessing the risk of bias of included studies. To address this limitation, we contacted the primary authors of those articles with missing data.

Although large numbers of women have been studied, Asia and Oceania had little
representation in the final selection of studies, with only one article from each region.
Moreover, some of the included studies had a relatively small sample size.

One limitation of our review was the high heterogeneity found when the meta-analysis combined studies reporting means and those reporting only mean differences. To compensate for this limitation, we presented a separate meta-analysis for those studies reporting only a mean difference. The main strength of this review is that, by restricting the population's inclusion criteria, we could control for the large heterogeneity between studies and calculate a reliable summary estimate that quantifies the bias that occurs when using self-reported weight and height data for women in reproductive age.

Finally, we observed that there is a difference in relation to the degree of significance in some analyzes. In this regard, the limited number of studies in some sub analysis challenged the interpretation of the results.

381 *Conclusions*

This review presents the difference of using self-reported weight and height compared to direct measurements in women of reproductive age with no eating disorders or conditions that may confound the comparison. The population selected in this study allowed us to reduce the heterogeneity between studies and to achieve a summary estimate of possible bias.

Self-reported maternal weight and height are broadly used, particularly in situations 387 where even basic anthropometric measurements cannot be taken. Self-reported 388 measures are used in clinical practice and in studies that relate them with pregnancy 389 outcomes. This review shows a low bias in the estimation of weight and height using 390 391 self-reported measures; for example, the BMI of a woman with a weight of 50kg and a height of 1.65mts, would differ by 2.36% (95%CI: 2.07%, 2.58%) if measured using 392 self-reported data. Our interpretation is that self-reported weight and height in women 393 of reproductive age is a measure that closely estimates the real values and can be used 394 as proxy of real values both in clinical and research evaluation. 395

Funding: The study was funded by Institute for Clinical Effectiveness and Health
Policy (IECS), Buenos Aires, Argentina. The funding source had no role in study
design, data collection, data analysis, data interpretation or preparation of the
manuscript.

400

401 References

Obesity: preventing and managing the global epidemic. Report of a WHO
 Consultation. Geneva: 2000.

404 2. McAdams MA, Van Dam RM, Hu FB. Comparison of self-reported and
405 measured BMI as correlates of disease markers in US adults. Obesity. 2007;15(1):188406 96.

Hanson MA, Bardsley A, De-Regil LM, Moore SE, Oken E, Poston L, et al. The
International Federation of Gynecology and Obstetrics (FIGO) recommendations on
adolescent, preconception, and maternal nutrition: "Think Nutrition First". Int J
Gynaecol Obstet. 2015;131 Suppl 4:S213-53.

4. 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. Obesity Rev. 2007;8(4):307-26.

414 5. Weaver EH, Gibbons L, Belizan JM, Althabe F. The increasing trend in preterm
415 birth in public hospitals in northern Argentina. IntJ Gynaecol Obstet. 2015;130(2):137416 41.

417 6. Shin D, Chung H, Weatherspoon L, Song WO. Validity of prepregnancy weight
418 status estimated from self-reported height and weight. Matern Child Health J.
419 2014;18(7):1667-74.

420 7. Rubinstein F, Micone P, Bonotti A, Wainer V, Schwarcz A, Augustovski F, et

al. Influenza A/H1N1 MF59 adjuvanted vaccine in pregnant women and adverse
perinatal outcomes: multicentre study. BMJ. 2013;346:f393.

423 8. Acevedo P, Lopez-Ejeda N, Alferez-Garcia I, Martinez-Alvarez JR, Villarino A,
424 Cabanas MD, et al. Body mass index through self-reported data and body image

425 perception in Spanish adults attending dietary consultation. Nutrition. 2014;30(6):679426 84.

427 9. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items
428 for systematic reviews and meta-analyses: the PRISMA statement. Ann Internal Med.
429 2009;151(4):264-9, W64.

430 10. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al.

The PRISMA statement for reporting systematic reviews and meta-analyses of studies

that evaluate healthcare interventions: explanation and elaboration. BMJ.

433 2009;339:b2700.

434 11. Bardach A, Ciapponi A, Garcia-Marti S, Glujovsky D, Mazzoni A, Fayad A, et
435 al. Epidemiology of acute otitis media in children of Latin America and the Caribbean:
436 A systematic review and meta-analysis. Int J Pediatr Otorhinolaryngol.

437 2011;75(9):1062-70.

438 12. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock

439 SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology

440 (STROBE): explanation and elaboration. PLoS Med. 2007;4(10):e297.

441 13. Review Manager (RevMan) 5.2 ed. Copenhagen: The Nordic Cochrane Centre;442 2012.

14. Landsberg B, Bastian I, Plachta-Danielzik S, Lange D, Johannsen M, Seiberl J,

et al. Self-reported and measured height and weight in adolescents. Gesundheitswesen
2011;73(1):40-5.

446 15. Avila-Funes JA G-RL, Ponce De Leon Rosales S. Validity of height and weight

self-report in Mexican adults: results from the national health and aging study.J

448 Nutrition Health Aging. 2004;8(5):355-61.

- Brettschneider AK, Rosario AS, Ellert U. Validity and predictors of BMI
 derived from self-reported height and weight among 11- to 17-year-old German
 adolescents from the KiGGS study. BMC Res Notes. 2011:414.
- 452 17. Brunner Huber LR. Validity of self-reported height and weight in women of
 453 reproductive age. Matern Child Health J. 2007:137-44.
- 18. Carvalho AM, Piovezan LG, Selem SS, Fisberg RM, Marchioni DM. Validation
 and calibration of self-reported weight and height from individuals in the city of Sao
 Paulo. Rev Bras Epidemiol. 2014:735-46.
- 457 19. Ekstrom S, Kull I, Nilsson S, Bergstrom A. Web-based self-reported height,
 458 weight, and body mass index among Swedish adolescents: a validation study. J Med
 459 Internet Res. 2015:e73.
- Elgar FJ, Roberts C, Tudor-Smith C, Moore L. Validity of self-reported height
 and weight and predictors of bias in adolescents. J Adoles Health. 2005:371-5.
- 462 21. Fonseca H, Silva AM, Matos MG, Esteves I, Costa P, Guerra A, et al. Validity
 463 of BMI based on self-reported weight and height in adolescents. Acta Paediatr.
- 464 2010:83-8.
- 465 22. Galan I, Gandarillas A, Febrel C, Meseguer C. [Validation of self-reported
 466 weight and height in an adolescent population]. Gaceta sanitaria / SESPAS.
- 467 2001;15(6):490-7.
- 468 23. Himes JH, Faricy A. Validity and reliability of self-reported stature and weight
 469 of US adolescents. Am J Hum Biol. 2001:255-60.
- 470 24. Larsen JK, Ouwens M, Engels RC, Eisinga R, van Strien T. Validity of self-
- 471 reported weight and height and predictors of weight bias in female college students.
- 472 Appetite. 2008:386-9.
- 473 25. Leatherdale ST, Laxer RE. Reliability and validity of the weight status and
- dietary intake measures in the COMPASS questionnaire: are the self-reported measures
- of body mass index (BMI) and Canada's food guide servings robust? Int J
- 476 Behav Nutr Phys Act. 2013:42.
- 477 26. Legleye S, Beck F, Spilka S, Chau N. Correction of body-mass index using
- 478 body-shape perception and socioeconomic status in adolescent self-report surveys.
- 479 PLoS One. 2014:e96768.
- Lo WS, Ho SY, Wong BY, Mak KK, Lam TH. Validity and test-retest reliability
 in assessing current body size with figure drawings in Chinese adolescents. Int J Pediatr
 Obes. 2011:e107-13.

483 28. Marrodan MD, Martinez-Alvarez JR, Villarino A, Alferez-Garcia I, Gonzalez484 Montero de Espinosa M, Lopez-Ejeda N, et al. Utility of self-reported anthropometric

data for evaluation of obesity in the Spanish population; study EPINUT-

486 ARKOPHARMA. Nutrición Hospitalaria. 2013:676-82.

Peixoto MdRG, D' Aquino Benicio MHD, Veiga Jardim PC. Validity of selfreported weight and height: the Goiânia study, Brazil. Rev Saude Publica 2006:106572.

490 30. Pregnolato TdS, Mesquita LM, Ferreira PG, Santos MM, Santos C, Carminhoto,

491 Costa RF. Validade de medidas autorreferidas de massa e estatura e seu impacto na

492 estimativa do estado nutricional pelo índice de massa corporal. Rev Bras Crescimento
493 Desenvolv Hum. 2009:35-41.

494 31. Pursey K, Burrows TL, Stanwell P, Collins CE. How accurate is web-based self495 reported height, weight, and body mass index in young adults? J Med Internet Res.
496 2014:e4.

497 32. Rodrigues PR G-SR, Pereira RA. Validity of self-reported weight and stature in
498 adolescents from Cuiabá, Central-Western Brazil. Rev Nutr. 2013;26(3):283-90.

33. Savane FR, Navarrete-Munoz EM, Garcia de la Hera M, Gimenez-Monzo D,
Gonzalez-Palacios S, Valera-Gran D, et al. [Validation of self-reported weight and
height university population and factors associated with differences between self

reported and measured antropometrics]. Nutrición hospitalaria. 2013:1633-8.

503 34. Shin D, Chung H, Weatherspoon L, Song WO. Validity of prepregnancy weight
status estimated from self-reported height and weight. Matern Child Health J.
2014:1667-74.

506 35. Unikel-Santoncini C, Ocampo-Ortega R, Zambrano-Ruiz J. Exactitud del
507 autorreporte de peso y talla en mujeres de 15 a 19 años del Estado de México. Salud
508 Publica Mex 2009:194-201.

509 36. Vitale R, Lavin Fueyo J, Rivera C, Mamondi V, Berra S. Validez del peso y la
510 talla declarados en adolescentes escolarizados de la ciudad de Córdoba. Rev salud
511 pública (Córdoba). 2013:42-8.

512 37. Gorber SC, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self513 report measures for assessing height, weight and body mass index: a systematic review.
514 Obesity Reviews. 2007;8(4):307-26.

18

515	38. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-
516	reported height and weight in women: an integrative review of the literature. Journal of
517	midwifery & women's health. 2003;48(5):338-45.
518	
519	
520	
521	
522	Legends
523	Figure 1. Flow Chart of screening and selection of studies.
524	Figure 2. Forest plot of self-reported vs. direct measured weight (panel A) and height
525	(panel B) (mean difference), by age group.
526	Figure 3. Forest plot of self-reported vs. direct measured weight (panel A) and height
527	(panel B) (mean difference), by region.
528	Figure 4. Forest plot of self-reported vs. direct measured weight (panel A) and height
529	(panel B) (mean difference), by time between self-reported and direct measured.
530	Figure 5. Forest plot of self-reported vs. direct measured weight (panel A) and height
531	(panel B) (mean difference), by Body Mass Index (BMI).
532	
533	Appendix S1. Supplementary Methods: MEDLINE search strategy. We include
534	search terms (Mesh and others) and description of how they were combined.
535	Table S1. Assessment of risk of bias by article. The findings of the present study that
536	the risk of bias assessment found that there was a high risk of bias in the selection of
537	participants in six studies and in the control of confounder in two.
538	Figure S1. Forest plot of self-reported vs. direct measured weight (panel A) and height
539	(panel B) (mean difference), by self-report method. The findings study suggested that
540	there was a negative difference if the information is gathered in through a long- distance

541 survey (-1.46kg [95%CI; -2.27, -0.64]; p=0.0004; I2=0%) or in a self-administered

questionnaire on-site (-1.14kg [95%CI -1.79, -0.48]; p=0.006; I2=54%) for weight. For
Height a significant difference if the information is gathered through a long- distance
survey (0.55cm [95%CI 0.00, 1.09]; p=0.05; I2=0%) or in an in-person interview
(0.65cm [95%CI; 0.28, 1.02]; p=0.0005; I2=38%). No important difference was found
when the data is gathered through an on-site self-administered questionnaire.

Figure S2. Forest plot of mean difference between self-reported and direct measured weight (panel A) and height (panel B) in studies that only reported mean differences. The study founding two studies only reported mean difference between methods, without specifying mean self-reported weight and mean direct measured weight. We found that, self-reported height was higher than direct measured height

Accepted

Table 1. Characteristics of included studies.

							<u> </u>		
First Author - Year	Country	Type of study	Population characteristic	Database analysed	Age range or mean <u>+</u> SD	Sample size	Reported outcome	SR* method	Time Lag between SR* and DM**
Brettschneider 2011 ⁽¹⁶⁾	Germany	Cross sectional	General population	KiGGS ⁵	14-17	948	MW,MH,MBMI,DW,DH,DBMI	Personal interview	Same day
Brunner 2007 ⁽¹⁷⁾	USA	Cross sectional	General population	CHIC ⁷	18-25 26-35 36-49	89 138 48	MW,MH,MBMI,DW,DH,DBMI	Self- administered survey	Same day
Carvalho 2014 ⁽¹⁸⁾	Brazil	Cross sectional	General population	ISA-Capital	12-19	32	MW,MH,MBMI	Long distance survey (Telephonic)	Different Days (non- specified)
Ekstrom 2015 ⁽¹⁹⁾	Stocolm	Prospective cohort	General population	-	16.5+0.4	889	MW,MH,MBMI,DW,DH,DBMI	Long distance survey (Online)	NR
Elgar 2005 ⁽²⁰⁾	Wale	Cross sectional	High school students	HBSC Study ⁴	15-17	211	MW,MH,MBMI,DW,DH,DBMI	Self- administered survey	Same day

Fonseca 2009 ⁽²¹⁾	Portugal	Cross sectional	High school students	HBSC Study ⁴	14 +1.8	233	MW,MH,MBMI,DW,DH,DBMI	Self- administered survey	Same day
Galán 2001 ⁽²²⁾	Spain	Cross sectional	High school students	-	15-18	1810	DW,DH,DBMI	Self- administered survey	Same da
Himes 2001 ⁽²³⁾	USA	Cross sectional	General population	NHANES III ³	12-16	876	MW,MH,MBMI,DW,DH,DBMI	NR	Same da
Larsen 2007 ⁽²⁴⁾	Netherlands	Cross sectional	University students	-	20.9+2.40	209	MW,MH,MBMI	Questionnaire on-site	Same da
Leatherdale 2013 ⁽²⁵⁾	Canada	Cross sectional	High school students	-	14-15	65	MW,MH,MBMI,DW,DH,DBMI	Self- administered survey	Differen Days (1 week
Legleye 2014 ⁽²⁶⁾	France	Cross sectional	General population	ESCAPAD ²	17-18	140	MW,MH,MBMI,DW,DH,DBMI	Self- administered survey	Same da
			×	C	13.67+1.18	1838			
Lo 2011 ⁽²⁷⁾	China	Cross sectional	High school students	HKSOS project ¹	16.29 +0.98	1275	MW,MH, MBMI	Questionnaire on-site	NR
Marrodan 2013 ⁽²⁸⁾	Spain	Cross sectional	General population	-	18-24 25-34	181 1486	MW,MH, MBMI	Personal interview	Same da

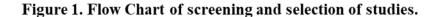
Peixoto 2006 ⁽²⁹⁾	Brazil	Cross sectional	General population	-	35-44 20-24 25-34 35-44	1876 97 184 150	DW,DH,DBMI	Personal interview	Same day
Pregnolato 2009 ⁽³⁰⁾	Brazil	Cross sectional	University students	- (28.3+11	549	MW,MH, MBMI	NR	Same day
Pursey 2014 ⁽³¹⁾	Australia	Cross sectional	General population	ed	18-35	93	MW, MH	Long distance survey (Online)	Different Days (<1 month)
Rodrigues 2013 ⁽³²⁾	Brazil	Cross sectional	High school students	-	14-19	40	MW,MH,MBMI,DW,DH,DBMI	NR	Same day
Savane 2013 ⁽³³⁾	Spain	Prospective cohort	University students	-	18 - 37	476	MW,MH,MBMI,DW,DH,DBMI	Self- administered survey	NR
Shin 2014 ⁽⁶⁾	USA	Cross	General	NHANES ⁸	16-25	1252	MW	Self-	Same day

		sectional	population		26-35	592		administered survey	
				_	36-44	599			
Unikel Santocini 2009 ⁽³⁵⁾	Mexico	Cross sectional	High school students	-	15-19	2357	MW,MH,DW,DH	Self- administered survey	Same day
Vitale 2013 ⁽³⁶⁾	Argentina	Cross sectional	High school students	-	15-18	61	MW,MH,MBMI	Self- administered survey	Same day

* self-reported ** directly measured. 1. Hong Kong Student Obesity Surveillance (HKSOS) project 2. ESCAPAD survey (Survey on health and behavior) 3. National Health and Nutrition Examination Survey III 4. Health Behavior School-Aged Children (HBSC) 5. German Health Interview and Examination Survey for Children and Adolescents (KiGGS) 6. Health Survey of São Paulo (ISA-Capital) 7. The Contraceptive History, Initiation, and Choice (CHIC) Study 8. National Health and Nutrition Examination. NR: not reported. MW: Measure Weight; MH: Measure Height; MBMI: Measure Body Mass Index; DW: Difference Weight; DH: Difference Height; DBMI: Difference Body Mass Index. SR: Self-reported; DM: Direct Measured. NR: Not reported.

Accept

Accepted manuscript



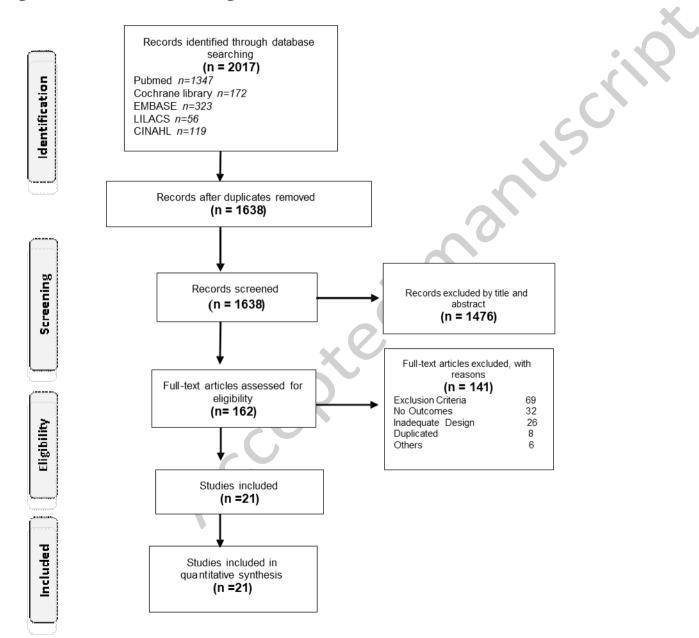


Figure 2. Forest plot of self-reported vs. direct measured weight (panel A) and height (panel B) (mean difference), by age group.

	Self-	reporte	bd	N	leasure	d		Mean D	ifference		Mear	Difference			Self	Reporte	ed	M	easured			Mean Difference	Mean	Difference
udy or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Rand	lom, 95% C		IV, Rai	dom, 95% CI		Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	IV, Fixe	ed, 95% CI
1.1 Age: 12 to 18 years														2.1.1 Age: 12 to 18 years										
ettschneider 2011		11.2	948		11.57	948	5.2%		[-1.73, 0.33]		-	•+		Brettschneider 2011	166	6.84	948	164.8	6.38	948	6.8%	1.20 [0.60, 1.80]		
arvalho 2014	56.75	11.54	32		11.03	32	0.2%	-0.42	[-5.95, 5.11]			-		Carvalho 2014		11.31	32		11.54	32	0.1%			
strom 2015	59	8.7	889	60.5		889	7.9%	-1.50	2.33, -0.67		_	-		Ekstrom 2015	167.9		889		6.2	889	7.4%			L
gar 2005	56.54				11.81	211	1.3%		[-3.37, 0.81]			+		Elgar 2005	162.93			162.38		211	1.4%			
inseca 2010		10.9	233	53.8		233	1.4%		[-3.17, 0.77]			+												
mes 2001		15.9	876			876	2.1%		[-2.82, 0.42]			+		Fonseca 2010	158.4	8.9	233		6.9	233		1.00 [-0.45, 2.45]		
atherdale 2013	57.55		65		11.35	65	0.4%		[-6.55, 1.05]			-		Himes 2001	160.9	10.6	876		8	876		-1.10 [-1.98, -0.22]		-
gleye 2014	57.39				9.72	140	1.1%		4.55, -0.17]			-		Leatherdale 2013	163.5			163.69		65		-0.19 [-2.72, 2.34]		
2011	45.5	7.7	1838	46.3	8	1838	21.3%		-1.31, -0.29			-		Legleye 2014	166.12	6.38	140	164.86	6.3	140	1.1%	1.26 [-0.23, 2.75]		<u> </u>
2011	48.9	6.9	1275			1275	17.8%		1.46, -0.34			-		Lo 2011	156.4	6.1	1838	156.4	6.3	1838	15.0%	0.00 [-0.40, 0.40]		+
odrigues 2013	54.5	7.5	40	54.3		40	0.5%		[-3.09, 3.49]			<u> </u>		Lo 2011	159.3	5.2	1275	159.2	5.8	1275	13.2%	0.10 [-0.33, 0.53]		+
nikel-Santoncini 2009	54.9	9.28	2032	56.24	10.54	2032	14.7%	-1.34	1.95, -0.73		-	-		Rodrigues 2013	162	10	40	161	10	40	0.1%	1.00 [-3.38, 5.38]		
tale 2013	54.05	9.02		54.82	7.82	45	0.4%		-4.30, 2.76					Unikel-Santoncini 2009	155.48	8.31	2012	155.34	6.02	2012	12.0%			+
ibtotal (95% CI)			8622			8624	74.4%	-1.05 [1.32, -0.78]	1				Vitale 2013	159		33		6.05	46	0.3%			
eterogeneity: Tau ² = 0.00				= 0.89); I ² = 0%									Subtotal (95% CI)	100	0.01	8592		0.00			0.24 [0.04, 0.44]		•
est for overall effect: Z = 7	7.59 (P <	0.0000	1)											Heterogeneity: Chi ² = 25.8	7 df = 12	/P = 0.0				0000	OLIO / F	0.2.1 [0.0.1, 0.1.1]		ľ
														Test for overall effect: Z = 2			, i –	34 90						
1.2 Age: 19 to 35 years														Test for overall effect. $Z = 2$	2.39 (P =	0.02)								
unner 2007	74.43				22.79	138	0.2%		[-7.20, 3.30]															
unner 2007	68.4					89	0.1%		[-7.80, 4.62]					2.1.2 Age: 19 to 35 years										
irsen 2008	62.3	8.2	203	65.7	9	203	2.0%		-5.07, -1.73					Brunner 2007	165.1	7.2		164.59		89	0.6%		_	
arrodan 2013	72.4				13.88	1486	5.5%		[-1.05, 0.95]			-		Brunner 2007	164.85			164.85		138	0.8%			<u> </u>
arrodan 2013	70.07				13.67	181	0.7%		[-3.71, 1.81]					Larsen 2008	170.8	6.3	203		6.5	203	1.6%	1.30 [0.05, 2.55]		<u> </u>
egnolato 2009	60.8		549			549	3.6%		-1.84, 0.64		-	-		Marrodan 2013	162.97	6.72	1486	162.56	6.77	1486	10.3%	0.41 [-0.07, 0.89]		+-
irsey 2014	65.82			66.48		93	0.2%		-5.76, 4.44			-		Marrodan 2013	163.41	6.74	181	162.9	6.71	181	1.3%	0.51 [-0.88, 1.90]	-	+
avane 2013	58.7		476		17.45		1.1%		-2.22, 2.22		_			Pregnolato 2009	163.1	6.6	549	162.1	6.5	549	4.0%	1.00 (0.23, 1.77)		
nin 2014	73		592		26.76	592	0.6%		[-4.31, 1.51]					Pursey 2014	166.5	6.56	93	165.2	6.21	93	0.7%	1.30 [-0.54, 3.14]		<u>+</u>
nin 2014 Ibtotal (95% CI)	65.9	28.31	1252	67.4	28.31	1252 5059	1.1%		[-3.72, 0.72] -1.86, -0.21]					Savane 2013	164.6	6	476			476	4.2%			<u>+-</u>
							15.2%	-1.04 [-1.80, -0.21]	1				Subtotal (95% CI)	101.0	•	3215					0.57 [0.25, 0.89]		•
eterogeneity: Tau ² = 0.46			f= 9 (P	= 0.17); l* = 30'	%								Heterogeneity: Chi ² = 4.42	df = 7 /0	= 0.72								1.
st for overall effect: Z = 2	2.47 (P =	0.01)												Test for overall effect: Z = 3			,1 - 0 9	0						
1.3 Age: 36 to 49 years														rest for overall effect. Z = 3	5.49 (P =	0.0005)								
unner 2007	77.97	15.00	40	04.40	18.78	48	0.1%	2.227	10.09, 3.65					2.1.3 Age: 36 to 49 years										
arrodan 2013	73.6		1876		12.58	1876	8.6%		10.09, 3.65															
radis 2008	69.98				16.62	372	8.0%		[-0.90, 0.70] [-2.90, 1.64]					Brunner 2007	164.85			164.85		48	0.3%			
in 2014	74.6		599		29.37	599	1.1%		[-2.90, 1.64] [-4.59, 1.79]					Marrodan 2013	161.9	6.41		161.39	6.41		14.3%			-
btotal (95% CI)	74.0	20.92	2895	/6	29.37	2895			[-4.59, 1.79]					Subtotal (95% CI)			1924			1924	14.6%	0.50 [0.09, 0.91]		•
eterogeneity: Tau ² = 0.00	Chiller 1	10 40		0.000	12 - 0.04	2000	10.07	-0.20	[-0.55, 0.47]			T		Heterogeneity: Chi ² = 0.13), df = 1 (F	P = 0.72)	; I ² = 0%	6						
st for overall effect: Z = 0			- 5 (P -	0.09),	1 = 0.96									Test for overall effect: Z = 2	2.41 (P =	0.02)								
st for overall effect: $Z = 0$	J.69 (P =	0.49)																						
tal (95% CI)			16576			16579	100.0%	0.04 [1.170.71			•		Total (95% CI)			13731			13744	100.0%	0.36 [0.20, 0.51]		•
eterogeneity: Tau ² = 0.00	0.058-1			n - 0 6	20. IZ - OI		100.0%	-0.94 [_		•		Heterogeneity: Chi ² = 33.9	6 df = 22	P = 0.0	15) I ² =	35%						
eterogeneity: Tau* = 0.00 est for overall effect: Z = 7				F = 0.5	3), 12 = 01	70				-10	-5	ó é	10	Test for overall effect: Z = 4									-4 -2	0 2 4
st for overall effect: Z = / st for subgroup differen				m - 0.4	2) IZ - 6	0.400					Self Report	d Direct Mea	sure	Test for subgroup differen				0 - 0 1 7	8-426	<i>~</i>			Favours [experimental	 Favours [control]
cioi subdroup differen	ces, uni*	= 4.03	ut = 2	$v_{r} = 0.1$	37, 11 = 5	0.4%								restion subditionbigueten	ues. Uni"	- 3.35,	ui = 2 (P	= 0.17).	1 = 43.0	20				

Panel A: self-reported vs. direct measured weight

Panel B: self-reported vs. direct measured height



Figure 3. Forest plot of self-reported vs. direct measured weight (panel A) and height (panel B) (mean difference), by region.

		f-report			easure			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI
1.2.4 Region: Latin Amer	rica & Ca	ribean							
Carvalho 2014	56.75	11.54	32	57.17	11.03	32	0.4%	-0.42 [-5.95, 5.11]	
Pregnolato 2009	60.8	10.3	549	61.4	10.7	549	6.7%	-0.60 [-1.84, 0.64]	-+
Rodrigues 2013	54.5	7.5	40	54.3	7.5	40	1.1%	0.20 [-3.09, 3.49]	
Jnikel-Santoncini 2009	54.9	9.28	2032	56.24	10.54	2032	19.5%	-1.34 [-1.95, -0.73]	+
/itale 2013 Subtotal (95% CI)	54.05	9.02	43 2696	54.82	7.82	45 2698	0.9%	-0.77 [-4.30, 2.76] -1.14 [-1.67, -0.61]	
	0.017	1.00 -		0.701	8-00	2030	20.0%	-1.14[-1.07, -0.01]	•
Heterogeneity: Tau ² = 0.0				U.76);	I* = U%				
Test for overall effect: Z =	4.21 (P	< 0.0001	0						
1.2.5 Region: Europe									
Brettschneider 2011	59	11.2	948	59.7	11.57	948	9.2%	-0.70 [-1.73, 0.33]	
Ekstrom 2015	59	8.7	889	60.5	9.2	889	12.8%	-1.50 [-2.33, -0.67]	+
Elgar 2005		10.02		57.82		211	2.6%	-1.28 [-3.37, 0.81]	+
onseca 2010	52.6	10.9	233	53.8	10.8	233	2.9%	-1.20 [-3.17, 0.77]	
arsen 2008	62.3	8.2	203	65.7	9	203	3.9%	-3.40 [-5.07, -1.73]	
_ealeye 2014	57.39	8.95		59.72	9.72	140	2.4%	-2.33 [-4.52, -0.14]	
Marrodan 2013		13.14		71.02		181	1.5%	-0.95 [-3.71, 1.81]	
Marrodan 2013		12.37	1876		12.58	1876	13.6%	-0.10 [-0.90, 0.70]	+
Marrodan 2013	72.4	13.8	1486	72.45		1486	9.7%	-0.05 [-1.05, 0.95]	+
Savane 2013		17.45	476		17.45	476	2.3%	0.00 [-2.22, 2.22]	
Subtotal (95% CI)			6643			6643	60.7%	-1.02 [-1.68, -0.37]	•
Heterogeneity: Tau ² = 0.5	i3; Chi ² =	19.81,	df = 9 (P	= 0.02)	; I ² = 559	6			
Fest for overall effect: Z =	3.05 (P =	= 0.002)		,					
1.2.6 Region: North Ame	rica								
Brunner 2007		21.73	138	76.38	22.79	138	0.4%	-1.95 [-7.20, 3.30]	
Brunner 2007		15.38		81.19		48	0.2%	-3.22 [-10.09, 3.65]	
Brunner 2007		20.57		69.99	21.7	89	0.3%	-1.59 [-7.80, 4.62]	
Himes 2001	58.9	15.9	876	60.1	18.6	876	4.1%	-1.20 [-2.82, 0.42]	
eatherdale 2013		10.75	65		11.35	65	0.8%	-2.75 [-6.55, 1.05]	
Shin 2014		24.33	592		26.76	592	1.4%	-1.40 [-4.31, 1.51]	
3hin 2014		28.31	1252		28.31	1252	2.3%	-1.50 [-3.72, 0.72]	
Shin 2014		26.92	599		29.37	599	1.1%	-1.40 [-4.59, 1.79]	
Subtotal (95% CI)			3659		24.41	3659	10.7%	-1.51 [-2.53, -0.48]	◆
Heterogeneity: Tau ² = 0.0	0: Chi ² =	0.83. dt	f= 7 (P =	1.00):	P = 0%				
Test for overall effect Z =									
Fotal (95% CI)			12998				100.0%	-1.01 [-1.36, -0.67]	•
Heterogeneity: Tau ² = 0.0				P = 0.34	 I² = 9⁶ 	%			-10 -5 0 5 10
Fest for overall effect: Z =	5.78 (P	< 0.0000	01)						Favours [experimental] Favours [control]

2 CC

Panel A: self-reported vs. direct measured weight

		Reporte			easured			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI
2.2.1 Region: Latin Amer									
Carvalho 2014		11.31	32		11.54	32	0.2%	-2.00 [-7.60, 3.60]	
Pregnolato 2009	163.1	6.6	549	162.1	6.5	549	7.2%	1.00 [0.23, 1.77]	
Rodrigues 2013	162	10	40	161	10	40	0.3%	1.00 [-3.38, 5.38]	
Unikel-Santoncini 2009	155.48	8.31		155.34	6.02	2012	12.4%	0.14 [-0.31, 0.59]	+
Vitale 2013 Subtotal (95% CI)	159	6.91	33 2666	158	6.05	46 2679	0.8%	1.00 [-1.94, 3.94] 0.43 [-0.07, 0.92]	
		10 46-		0.241-17-	4400	2013	20.3%	0.45 [-0.07, 0.52]	
Heterogeneity: Tau ² = 0.0			: 4 (P =	U.34); I* =	11%				
Test for overall effect: Z =	1.70 (P =	0.09)							
2.2.2 Region: Europe									
Brettschneider 2011	166	6.84	990	164.8	6.38	990	9.9%	1.20 [0.62, 1.78]	
Ekstrom 2015	167.9	6.1	889	167.4	6.2	889	10.1%	0.50 [-0.07, 1.07]	+
Elgar 2005	162.93	7.42	221	162.38	6.03	221	3.5%	0.55 [-0.71, 1.81]	
Fonseca 2010	158.4	8.9	233	157.4	6.9	233	2.8%	1.00 [-0.45, 2.45]	<u>+</u>
Larsen 2008	170.8	6.3	209	169.5	6.5	209	3.7%	1.30 [0.07, 2.53]	
Legleye 2014	166.12	6.38	140	164.68	6.3	140	2.7%	1.44 [-0.05, 2.93]	
Marrodan 2013	161.9	6.41	1876	161.39	6.41	1876	13.2%	0.51 [0.10, 0.92]	+
Marrodan 2013	162.97	6.72	1486	162.56	6.77	1486	11.7%	0.41 [-0.07, 0.89]	
Marrodan 2013	163.41	6.74	181	162.9	6,71	181	3.0%	0.51 [-0.88, 1.90]	
Savane 2013	164.6	6	476	164.3	6	476	7.3%	0.30 [-0.46, 1.06]	+
Subtotal (95% CI)			6701			6701	67.8%	0.63 [0.41, 0.85]	•
Heterogeneity: Tau ² = 0.0	0; Chi ² = 8	1.30, df =	9 (P =	0.50); l² =	0%				
Test for overall effect: Z =	5.59 (P <	0.00001)						
2.2.3 Region: North Ame	rica								
Brunner 2007	164.85	7.07	48	164.85	6.86	48	0.8%	0.00 (-2.79, 2.79)	
Brunner 2007	164.85	7.52		164.85	7.52	138	1.9%	0.00 [-1.77, 1.77]	
Brunner 2007	165.1	7.2	89	164.59	6.51	89	1.5%	0.51 [-1.51, 2.53]	
Himes 2001	160.9	10.6	876	162	8	876	6.1%	-1.10 [-1.98, -0.22]	
Leatherdale 2013	163.5	7.95	65	163.69	6.72	65	1.0%	-0.19 [-2.72, 2.34]	
Subtotal (95% CI)			1216			1216	11.4%	-0.62 [-1.30, 0.06]	•
Heterogeneity: Tau ² = 0.0			: 4 (P =	0.54); I² =	0%				
Test for overall effect: Z =	1.78 (P=	0.08)							
Total (95% CI)			10583				100.0%	0.49 [0.23, 0.75]	•
Heterogeneity: Tau ² = 0.0				= 0.08);	P= 32%				-4 -2 0 2 4
Test for overall effect: Z =									Favours [experimental] Favours [control]
Test for subgroup differer	nces: Chi²	= 11.68	df = 2	P = 0.00	3), I ² = 83	2.9%			r aroure [experimental] T aroure [control]

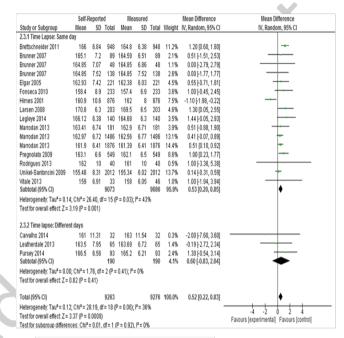
Panel B: self-reported vs. direct measured height

Figure 4. Forest plot of self-reported vs. direct measured weight (panel A) and height (panel B) (mean difference), by time between self-reported and direct measured.

		-Report			easured			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI
1.3.1 Time Lapse: Same	day								
Brettschneider 2011	59	11.2	948	59.7	11.57	948	10.7%	-0.70 [-1.73, 0.33]	
Brunner 2007	68.4	20.57	89	69.99	21.7	89	0.3%	-1.59 [-7.80, 4.62]	
Brunner 2007	77.97	15.38	48	81.19	18.78	48	0.3%	-3.22 [-10.09, 3.65]	
Brunner 2007	74.43	21.73	138	76.38	22.79	138	0.5%	-1.95 [-7.20, 3.30]	
Elgar 2005	56.54	10.02	211	57.82	11.81	211	2.8%	-1.28 [-3.37, 0.81]	+
Fonseca 2010	52.6	10.9	233	53.8	10.8	233	3.1%	-1.20 [-3.17, 0.77]	
Himes 2001	58.9	15.9	876	60.1	18.6	876	4.6%	-1.20 [-2.82, 0.42]	
Larsen 2008	62.3	8.2	203	65.7	9	203	4.3%	-3.40 [-5.07, -1.73]	<u> </u>
Legleye 2014	57.39	8.95	140	59.72	9.72	140	2.5%	-2.33 [-4.52, -0.14]	
Marrodan 2013	73.6	12.37	1876	73.7	12.58	1876	16.4%	-0.10 [-0.90, 0.70]	+
Marrodan 2013	70.07	13.14	181	71.02	13.67	181	1.6%	-0.95 [-3.71, 1.81]	
Marrodan 2013	72.4	13.8	1486	72.45	13.88	1486	11.2%	-0.05 [-1.05, 0.95]	+
Pregnolato 2009	60.8	10.3	549	61.4	10.7	549	7.5%	-0.60 [-1.84, 0.64]	-+
Rodrigues 2013	54.5	7.5	40	54.3	7.5	40	1.1%	0.20 [-3.09, 3.49]	
Shin 2014	74.6	26.92	599	76	29.37	599	1.2%	-1.40 [-4.59, 1.79]	
Shin 2014	65.9	28.31	1252	67.4	28.31	1252	2.5%	-1.50 [-3.72, 0.72]	+
Shin 2014	73	24.33	592	74.4	26.76	592	1.5%	-1.40 [-4.31, 1.51]	
Unikel-Santoncini 2009	54.9	9.28	2032	56.24	10.54	2032	25.2%	-1.34 [-1.95, -0.73]	+
Vitale 2013	54.05	9.02	43	54.82	7.82	45	1.0%	-0.77 [-4.30, 2.76]	
Subtotal (95% CI)			11536			11538	98.3%	-0.97 [-1.37, -0.57]	◆
Heterogeneity: Tau ² = 0.1	0; Chi ² =	21.07, 0	if = 18 (P = 0.28	3); I² = 15	%			
Test for overall effect: Z =	4.73 (P	< 0.0000	11)						
1.3.2 Time Lapse: Differe	ent days								
Carvalho 2014	56.75	11.54	32	57.17	11.03	32	0.4%	-0.42 [-5.95, 5.11]	
Leatherdale 2013	57.55	10.75	65	60.3	11.35	65	0.9%	-2.75 [-6.55, 1.05]	
Pursey 2014	65.82	17.7	93	66.48	17.8	93	0.5%	-0.66 [-5.76, 4.44]	
Subtotal (95% CI)			190			190	1.7%	-1.64 [-4.30, 1.03]	-
Heterogeneity: Tau ² = 0.0	0; Chi² =	0.66, df	= 2 (P =	0.72);	l² = 0%				
Test for overall effect: Z =	1.20 (P =	= 0.23)							
Total (95% CI)			11726			11728	100.0%	-0.96 [-1.31, -0.60]	•
Heterogeneity: Tau ² = 0.0	3; Chi ² =	21.99, (if = 21 (P = 0.40); ² = 49	6			-10 -5 0 5 1
Test for overall effect Z =	5.30 (P	< 0.0000	11)						Favours [experimental] Favours [control]
Test for subaroup differe	nrae [.] Ch	$i^2 = 0.24$	df = 1 i	P=06	3) I ² = 0	X.			Favours (experimental) Favours (Control)

Panel A: self-reported vs. direct measured weight

x cet



Panel B: self-reported vs. direct measured height

Figure 5. Forest plot of self-reported vs. direct measured weight (panel A) and height (panel B) (mean difference), by Boby Mass Index (BMI).

Mean Difference	Mean Difference		Mean Difference Mean Difference
Study or Subgroup Mean Difference SE Weight IV, Random, 95% CI	IV, Random, 95% CI		Study or Subgroup Mean Difference SE Weight IV, Random, 95% CI IV, Random, 95% CI
1.5.1 BMI: Low weight			2.5.1 BMI: Low weight
Peixoto 2006 0.81 0.3 10.5% 0.81 [0.22, 1.40]			Peixoto 2006 3.66 0.74 3.4% 3.66 [2.21, 5.11]
Marrodan 2013 -0.8 0.73 3.6% -0.80 [-2.23, 0.63]			Marrodan 2013 0.7 2.14 0.5% 0.70 [-3.49, 4.89]
Subtotal (95% CI) 14.1% 0.14 [-1.41, 1.70]			Subtotal (95% CI) 3.9% 2.86 [0.29, 5.44]
Heterogeneity: Tau ² = 0.98; Chi ² = 4.16, df = 1 (P = 0.04); I ² = 76%			Heterogeneity: Tau ² = 1.82; Chi ² = 1.71, df = 1 (P = 0.19); i ² = 41%
Test for overall effect: Z = 0.18 (P = 0.86)			Test for overall effect: Z = 2.18 (P = 0.03)
1.5.2 BMI: Normal Weight			2.5.2 BMI: Normal Weight
Peixoto 2006 0.27 0.1 16.1% 0.27 [0.07. 0.47]	-		Peixoto 2006 1.71 0.21 11.2% 1.71 [1.30, 2.12]
Marrodan 2013 -0.4 0.12 15.7% -0.40 [-0.64, -0.16]	-		Marrodan 2013 0.3 0.03 13.9% 0.3010.24.0.36
Subtotal (95% Cl) 31.8% -0.06 [-0.72, 0.59]	-		Subtotal (95% Cl) 25.1% 0.99 [-0.39, 2.37]
Heterogeneity: Tau ² = 0.21; Chi ² = 18.40, df = 1 (P < 0.0001); l ² = 95%			Heterogeneity: Tau ² = 0.97; Chi ² = 44.18, df = 1 (P < 0.00001); l ² = 98%
Test for overall effect: Z = 0.18 (P = 0.85)			Test for overall effect: Z = 1.40 (P = 0.16)
1.5.3 BMI: Overweight			2.5.3 BMI: Overweight
			Peixoto 2006 2.28 0.25 10.3% 2.28 [1.79, 2.77]
Peixoto 2006 -0.51 0.19 13.7% -0.51 [-0.88,-0.14] Marrodan 2013 -0.34 0.12 15.7% -0.34 [-0.58,-0.10]			Marrodan 2013 0.51 0.04 13.8% 0.51 [0.43, 0.59]
Subtotal (95% CI) 29.4% -0.39 [-0.59, -0.19]			Subtotal (95% Cl) 24.1% 1.38 [-0.36, 3.11]
Heterogeneity: Tau ² = 0.00; Chi ² = 0.57, df = 1 (P = 0.45); I ² = 0%	•		Heterogeneity: Tau ² = 1.53; Chi ² = 48.88, df = 1 (P < 0.00001); i ² = 98%
Test for overall effect: $Z = 3.83$ (P = 0.0001)			Test for overall effect: Z = 1.56 (P = 0.12)
1.5.4 BMI: Obese			2.5.4 BMI: Obese
Peixoto 2006 -0.52 0.34 9.5% -0.52 [-1.19, 0.15]			Peixoto 2006 3.4 0.41 7.2% 3.40 [2.60, 4.20]
Marrodan 2013 -1.01 0.86 0.0% -1.01 [-2.70, 0.68]			Marrodan 2013 0.73 0.08 13.5% 0.73 [0.57, 0.89]
Marrodan 2013 -0.48 0.57 5.2% -0.48 [-1.60, 0.64]			Marrodan 2013 0.95 0.14 12.6% 0.95 [0.68, 1.22]
Marrodan 2013 -0.28 0.32 10.0% -0.28 [-0.91, 0.35]			Marrodan 2013 0.97 0.06 13.7% 0.97 [0.85, 1.09]
Subtotal (95% CI) 24.7% -0.41 [-0.83, 0.02]	-		Subtotal (95% CI) 46.9% 1.24 [0.83, 1.65]
Heterogeneity: Tau ² = 0.00; Chi ² = 0.28, df = 2 (P = 0.87); l ² = 0%		·	Heterogeneity: Tau ² = 0.14; Chi ² = 42.97, df = 3 (P < 0.00001); i ² = 93%
Test for overall effect: Z = 1.88 (P = 0.06)			Test for overall effect: Z = 5.91 (P < 0.00001)
Total (95% CI) 100.0% -0.19 [-0.49, 0.12]	•		Total (95% CI) 100.0% 1.26 [0.95, 1.57]
Heterogeneity: Tau ² = 0.14; Chi ² = 40.70, df = 8 (P < 0.00001); I ² = 80%			Heterogeneity: Tau ² = 0.18; Chi ² = 275.42, df = 9 (P < 0.00001); i ² = 97%
Test (as suppl) effect: $7 = 4.24$ ($D = 0.22$)	-2 0 2 4 ours [experimental] Favours [control]		Test for overall effect: Z = 8.02 (P < 0.00001) -20 -10 Favours [experimental] Favours [control]
Test for subgroup differences: Chi ² = 1.32, df = 3 (P = 0.73), I ² = 0%	urs (experimental) - avous (control)		Test for subgroup differences: Chi ² = 1.66, df = 3 (P = 0.65), I ² = 0%



