Title: The Accuracy of Self-Reported Pregnancy-Related Weight: A Systematic Review

Authors: Irene Headen, PhD MS[a], Alison K Cohen, MPH[b], Mahasin Mujahid, PhD, MS[b], Barbara Abrams, DrPH, MPH[b]

Author Affiliations:
[a]University of California Berkeley, School of Public Health, Division of Community Health Science, 207 University Hall, Berkeley, CA, USA
[b]University of California Berkeley, School of Public Health, Division of Epidemiology, 101 Haviland Hall, Berkeley, CA, USA

Submitted version, published as: Headen I, Cohen AK, Mujahid M, Abrams B. "The accuracy of self-reported pregnancy-related weight: A systematic review." *Obesity Reviews*. 2017; 18(3): 350-369. doi: 10.1111/obr.12486

Keywords: Validity, weight, measurement, pregnancy, postpartum period

Running title: Accuracy of Self-Reported Pregnancy-Related Weight

Acknowledgements: This work was supported by funding from the National Institute of Minority Health and Health Disparities (grant R01MD006014) and the Health Resources and Service Administration (grant T76MC00002). We would like to acknowledge the contributions of Ruvani Jayaweera and Esther Chung for their assistance with locating and organizing papers.

Corresponding Author: Irene Headen, PhD
525 University Hall, University of California Berkeley, Berkeley, CA, 94720
iheaden@berkeley.edu; Phone: (919)-414-2267; Fax: (510)-642-9083

Conflicts of Interest: The authors have received funding from National Institute of Minority Health and Health Disparities and the Health Resources and Service Administration. Dr. Barbara Abrams is also a Reviewer for UpToDate Clinical Decision Support Resource. They have no other conflicts of interest to declare.

Abbreviations: GWG = Gestational weight gain; BMI= Body mass index; NCPP= National Collaborative Perinatal Project; SD = standard deviation; SES= socioeconomic status

ABSTRACT

Self-reported maternal weight is error-prone, and the context of pregnancy may impact error distributions. This systematic review seeks to summarize error in self-reported weights across pregnancy and assess implications for bias in associations between pregnancy-related weight and birth outcomes. We searched electronic databases (PubMed and Google Scholar) through November 2015 for peer-reviewed articles reporting accuracy of self-reported, pregnancy-related weight at four time points: prepregnancy, delivery, over gestation, and postpartum. Included studies compared maternal self-report to anthropometric measurement or medical report of these weights. Sixty-two studies met inclusion criteria. We extracted data on magnitude, direction, and variability of reporting error, and misclassification. We assessed impact of reporting error on bias in associations between pregnancy-related weight and birth outcomes. Women underreported prepregnancy (PPW: -2.94 to -0.29kg) and delivery weight (DW: -1.28 to 0.07kg), and over-reported gestational weight gain (GWG: 0.33 to 3kg). Magnitude of error was small, but ranged widely and varied by prepregnancy weight class and race/ethnicity. Misclassification was moderate (PPW: 0-48.3%; DW: 39-49%; GWG: 16.7-59.1%), and biased estimates of population prevalence, especially for excessive gestational weight gain. However, reporting error did not largely bias associations between pregnancy-related weight and birth outcomes. Though measured weight is preferable, self-report is a cost-effective and practical measurement approach. Future researchers should develop bias correction techniques for self-reported pregnancy-related weight.

INTRODUCTION

Existing literature has identified maternal weight as a key contributor to pregnancy and fetal health, as well as long-term health outcomes for mother and child (1-3). Valid evidence requires the best possible measurement of study variables. The gold standard for assessing body weight is measurement, by trained personnel, with a calibrated scale using standardized procedures to weigh subjects wearing light-weight clothing and no shoes (4). However, many studies of weight in childbearing women do not meet this rigorous standard, relying instead on self-reported weight (1,5). Though practical, using self-reported weight decreases precision and is prone to bias, limiting interpretation of study findings. The extent and impact of self-reported weight errors on study results in the context of pregnancy are poorly understood.

Maternal prepregnancy weight, gestational weight gain (GWG) and postpartum weight are relevant indicators for studies of maternal and child health. While they are all measured in slightly different ways, they all, in practice, incorporate maternal self-report to varying degrees. *Prepregnancy weight* is almost universally based on maternal recall, since few opportunities to measure women's weight prior to conception arise (1). Non-pregnant women of childbearing age underreport their weight from 0.2kg to 3.54kg (6-8), and correction techniques, such as regression calibration (9,10), have been developed to address this error (9,10). However, characteristics of reporting error for prepregnancy weight may differ because women often report this weight after conception. Recall predominantly occurs at first prenatal visit, but may be obtained any time during or after pregnancy, creating potential for different magnitudes of error. Many studies express maternal weight status using body mass index (BMI, weight(kg)/height(m)^2), which can conflate weight and height errors.

Typically, total *gestational weight gain* (amount of weight gained between conception and delivery) is anchored on *last weight prior to delivery,* which may be measured or reported when women are admitted for delivery, or may reflect measured weight taken at last prenatal visit prior to delivery (often within 1week but varies by gestational age (1,11)). However, many studies base gestational weight gain on self-reported total weight gain (1,5).

Postpartum weight retention (amount of weight retained compared to weight prior to conception) is anchored on *postpartum weight* evaluated within the year following delivery. This weight often reflects weight measured at a postpartum clinic visit or obtained by maternal report if a woman is recruited in to a study population. Such studies vary on the time after delivery at which they collect weight measurement and often do not collect repeated postpartum weight measures (12).

Self-reported measures at any of these times during pregnancy can introduce bias if associated error is differentially distributed within the population (13). Understanding magnitude, direction and impacts of such error on observed associations between pregnancy-related weight and birth outcomes is important and can lead to development of approaches to address bias for pregnancy-specific weight measures. Additionally, there is growing use of self-reported weight from US birth certificates in studies of maternal weight (e.g. (5,14-16)).

Two non-systematic reviews have investigated reporting error in pregnancy-related weight measures (9,17). One focuses on error surrounding different techniques used to measure weight during pregnancy (9). The other focuses only on reporting error in prepregnancy weight only (17). However, neither focused on implications of reporting error for misclassification or empirically assessed bias present in associations when self-reported-weight measures were used in analyses. Furthermore, a number of key studies have been published since these reviews were conducted (9,17). This indicates the need for an updated and systematic review.

In this study, we systematically review data from clinical interventions, trials, observational studies, and validation studies to assess accuracy of self-reported pregnancy-related weight. In addition to documenting magnitude and direction of reporting error, we focus on misclassification across clinically-relevant weight categories at four key time points: prepregnancy, delivery, duration of gestation, and postpartum. We then investigate variation in reporting error across maternal demographic and health characteristics. Next, we assess impact of observed reporting error on bias in estimated associations between pregnancy-related weight and birth outcomes. Finally, we discuss methodological limitations of current literature and offer recommendations to address bias in studies going forward.

METHODS

Search Strategy

We searched PubMed and Google Scholar search engines for relevant articles. Google Scholar was used to complement PubMed because it searches articles' full text for keywords, identifying articles that report on accuracy of self-reported pregnancy-related weight as supplementary analyses within the methods, results, or discussion sections. Such papers would have been overlooked in traditional PubMed searches. Google Scholar also includes a wider array of databases in searches (18),

Our search queries combined terms related to self-report, pregnancy, prepregnancy, gestational weight gain, delivery weight, validity, accuracy and recall. A full list of all queries executed is included in Supplemental Table1. In Google Scholar, we reviewed results from queries in the order provided (sorted by citation frequency and relevance; (18)) until 50 consecutive results failed to provide a relevant article. This technique mitigated the fact that many queries returned over 2,000 articles. Searches included literature published through November 25, 2015. We reviewed titles and abstracts to determine if studies met inclusion criteria, and conducted full text reviews of studies that did. We reviewed reference lists of selected articles to identify additional papers. Eight studies were identified through this method, three of which met all inclusion criteria.

Study Selection

Eligible studies compared maternal report of pregnancy-related weight (prepregnancy weight, GWG, delivery weight, and/or postpartum weight) to an anthropometric or medically reported measure of that weight. While medical reports of weight often rely on patient self-reports, we included them as acceptable weight references because they often reflect procedures conducted by trained personnel in the clinical setting (19), and are used as reference points for assessing accuracy of self-reported weight. All studies had to be published in English. Although self-reported height is a potential source of bias in self-reported BMI for pregnant women, we considered height measurement beyond the scope of this review and did not exclude studies based on method of height measurement. Self-reported height in non-pregnant populations is reported with error (6-8,20). Ideally, height would be measured during pregnancy, but height does not vary much over pregnancy, suggesting that errors are similar in non-pregnant and pregnant populations.

Data Extraction

We developed a data extraction form to identify relevant information on self-reported, medically recorded, and anthropometrically measured pregnancy-related weight at time points of interest. We extracted multiple measures of accuracy from each study. Both continuous and categorical measures of pregnancy-related weight are relevant for analytic purposes, and we extracted data on error for both variable types (21-23). Correlation coefficients are informative for assessing linear relationships between continuous weight measures, but do not provide adequate information on absolute error (23,24). Thus, we also extracted information on mean difference between self-reported and reference weight and variability of mean difference within study populations (21,23). For categorical weight measures, misclassification better quantifies error (22). We specifically focused on misclassification by clinically-relevant categories for prepregnancy weight (25) and GWG (1) because such categories are relevant for assessing risk of complications during pregnancy and determining whether women are meeting GWG recommendations. Overall, we extracted information on five measures of accuracy: mean difference between self-reported and reference weight, variability of mean difference, correlation coefficient or agreement statistic (i.e. kappa), percent agreement, and misclassification across relevant weight categories. We abstracted this information by demographic and health status subgroups to assess variation of reporting error across these characteristics.

We extracted information on source of weight reference (22,23). Gold standard weight references required use of measured weight at all times, with light clothing, no shoes, and appropriately calibrated equipment (4). Prepregnancy weight required measurement prior to conception (1,26). Delivery weight required measurement at time of delivery or at last clinical visit prior to delivery. Postpartum weight required measurement within one year after delivery. GWG required the use of gold standard weight measurements both for prepregnancy weight and delivery weight, as described above (1,26). Other methods of assessing reference weight, including measured weight at the wrong time (such as during the first trimester for prepregnancy weight) or use of medical records that included self-reported prepregnancy weight, were considered "alloyed" gold standards. Information on timing of reference weight measurement and length of recall was also collected.

Validity Assessment Rating

Because multiple measures of accuracy are needed to comprehensively quantify error for pregnancy-related weight (21-23), we scored studies based on their number of reported measures of accuracy. Studies received one point for each of the following measures: magnitude of error, variability of error, correlation coefficient/agreement statistic, and misclassification. Because BMI-based weight categories (25) and GWG adequacy categories (1) play a critical role in accurately directing prenatal care and assessing risk for adverse birth outcomes in clinical settings (1), bias due to misclassification across these categories can be particularly problematic. Thus, we gave two additional points to studies reporting misclassification across of these weight groups.

We also scored studies based on source of reference weight and sample size. Studies using gold standard reference weights were given three points because accuracy measures from such studies were least likely to be biased. Studies using alloyed gold standard measures were scored lower because of the potential for residual bias ((22,27,28). Studies using alloyed gold standards with measured weights taken at non-ideal times received two points. Studies using alloyed gold standards from medical records received one point because medical records are often still based on self-report within the clinical context (1,5,19). Finally, one point was added for studies with sample sizes ≥150 (27,29) as estimates of error from these studies may be more precise than those from smaller studies. There are many factors to consider when determining optimal sample size for validation studies (27,29,30); we did not find clear criteria on appropriate thresholds for larger sample sizes and so did not allot extra points for them. Overall, as detailed in Table 1, studies could receive 2 to 10 points based on how completely they assessed accuracy of self-reported pregnancy-related weight.

Study Synthesis

Use of self-reported pregnancy-related weight data is of concern because reporting error may bias measures of association when used to assess relationships with birth outcomes (22). Therefore, we evaluated measures of accuracy extracted from studies based on implications for this type of bias. We used both a qualitative and empirical approach. Qualitative assessment used the magnitude of each measure of accuracy to infer potential bias resulting from self-reported weight error. This approach assumed that higher error estimates (i.e. lower accuracy) indicated higher risk of bias and higher likelihood of bias substantial enough to change interpretation of associations (22,31). We used the set of criteria described in Table 2 to qualitatively classify risk of bias as low, moderate, or high. Since we were unable to find criteria designed specifically for pregnant populations, these criteria were based on those used to evaluate reporting error in non-pregnant populations (7,32), with modifications to define more explicit criteria for misclassification (33).

Empirical assessment of risk of bias was based on evidence from studies that explicitly compared differences in measures of association using self-reported and reference weight values. This approach provided explicit quantification of impacts of bias. Measures of association from these studies were evaluated based on whether self-reported weight changed the magnitude of association and the fundamental interpretation (i.e. significant or not) of the relationship (yes/no).

RESULTS

Figure 1 displays results from our study selection. PubMed searches returned 533 studies; Google Scholar searches returned 2,780-28,800 studies per search. Across all searches, 88 studies were identified for full-text review. After full-text review, 71 studies still met inclusion and exclusion criteria. Of these, a total of 18 studies reported on the same study populations; we selected one to report on going forward:

four used Project Viva (34-37); three used Kaiser Permanente Northern California (26,38,39); two used the Generation R study in the Netherlands (40,41); two used the Danish National Birth cohort (42,43); two used the Asthma Coalition on Community, Environment, and Social Stress cohort (44,45); two used the Mater-University of Queensland Study of Pregnancy (46,47); two used the National Collaborative Perinatal Project (NCPP) (48,49); and two studies used the NELLI cohort in Finland (50,51). For the NCPP population, we report findings for both studies because they report accuracy measures for different times during pregnancy. For Kaiser Permanente Northern California populations, Han et al. (39) is reported separately because they reports births from a different time frame than Ferrara et al. (38) and Hedderson et al. (26). Multiple studies were conducted at McGee Women's Hospital in Pittsburgh, PA (5,15,52,53), but populations included did not overlap. Overall, we abstracted data on reporting error from 62 study populations. Studies reported multiple pregnancy-related weight outcomes, resulting in a total of 54 studies on prepregnancy weight, 6 studies on delivery weight, 14 studies on GWG, and 1 study on postpartum weight. Detailed descriptions of studies can be found in Supplementary Table2.

Most studies (62.9%) were conducted in the United States. Over half (61.3%) of populations were racially and ethnically heterogeneous (5,14-16,19,35,38,39,41,44,49,52-78). Cohorts represented births from 1959 to 2013 and length of recall ranged from three days (20) to 32 years in the past (48,49). All but 18 studies (5,14,16,48,49,55,56,65,66,68,72,73,76,79-81) were based on convenience samples of women recruited from clinics or hospitals at the time of their first prenatal visit or at time of delivery. Six studies were based on national samples from the US

(48,49,54,62,66,76). Three studies were conducted in non-US, population-based samples; one in the Netherlands (41); one in Denmark (43); and one in the United Kingdom (80).

Validity Assessment Rating

Figure 2 displays the distribution of scores on our validity assessment scale. None of the 75 studies identified received a score of ten, and four (5.3%) received a score of nine. Twenty-nine studies (38.7%) scored five to eight points. Most studies (56.0%) received a score of four or lower, indicating that they did not report multiple measures of accuracy and/or relied on alloyed gold standards to assess validity.

Correlation coefficients or agreement statistics were the most widely reported measure of accuracy, with 47 studies (62.7%) reporting these measures (Supplementary Table3). Mean difference was reported in 43 studies (57.3%) and 18 also reported a measure of variability for this estimate. Misclassification was reported in 27 studies (36.0%), and 16 reported misclassification by clinically-relevant categories. True gold standard weight references were used in 16 studies (21.3%). Of those using alloyed gold standards, 35 (46.6%) used measured weight at a non-ideal time point and 24 (32.0%) used medical records. Most studies (n=56; 74.6%) had samples ≥150 (Supplementary Table3).

Studies scoring higher on our scale of completeness provide more comprehensive evidence on validity of self-reported pregnancy-related weight. Therefore, going forward, we report findings from 33 studies with that scored five or higher (44.0%). These studies were more likely to report two or more measures of accuracy and use a stronger reference weight measure.

Magnitude of Error in Self-Reported Weight Measures

Prepregnancy Weight

Twenty-three studies reporting validity of self-reported prepregnancy weight scored high on our validity assessment scale. Table 3 reports measures of accuracy; studies are ordered by length of recall (shortest to longest) and source of reference weight. Length of recall ranged from 9.2 weeks post conception (53) to 32 years in the past (48). A total of 16 studies had short lengths of recall (within 20 weeks of conception; (19,38,39,43,53,56-60,63,68,82-85)), five studies had medium lengths of recall (20 weeks after conception to 1 year post-pregnancy; (5,14,35,66,86)), and two studies had long lengths of recall (>1 year post-pregnancy; (48,72)). Studies with short and medium length recalls had similar findings and are reported together here. Nine of the 21 short/medium length recall studies used gold-standard weight references (35,38,39,57-60,84,85). Among these gold-standard studies, correlation between self-reported and measured weight was high (r=0.90(58)—0.99(35)). Women underreported their prepregnancy weight by 0.34kg(39) to 2.94kg (57). There variability around this mean difference was moderate (standard deviation (SD) range: 2.2kg(58)—5kg(59,84)). Correct classification of women into relevant weight classes was high (86.7(39)-91%(43)), but varied by prepregnancy weight class. Underweight (23.5% (39)), overweight (16.5(39) – 27.0%(57)), and obese class I (24.3 (39)) women had the highest prevalence of misclassification (Table 3).

Findings from studies with short/medium lengths of recall using alloyed gold standard weight references were similar to those using gold standard weight references, but suggested a higher prevalence of misclassification (Table 3). All studies with long lengths of recall used alloyed gold standards. Despite the longer time until recall, mean difference was only slightly smaller and correlation slightly lower compared to studies with short/medium lengths of recall.

Delivery Weight

Three studies reporting validity of self-reported delivery weight scored highly on our validity assessment (Table 2; (20,57,76)). Two of these studies had short lengths of recall (within six weeks of delivery; (20,57)). The remaining study had a long length of recall, ranging from 6-31 months post-pregnancy (76). All studies used gold standard weight references. Reporting error for studies with short lengths of recall ranged from -0.3kg(20) to 0.07kg ((57); Table 3). Misclassification, based on agreement within 1kg, was substantial (49%; (20)). The study with a longer length of recall reported greater mean difference and women were more likely to underestimate delivery weight. Prevalence of misclassification was lower, although misclassification was based on agreement within 2.27kg (5lb; (76)).

Gestational Weight Gain

Seven studies reporting validity of self-reported GWG scored highly on our validity assessment scale (Table 3). Of these, five had short lengths of recall that happened either shortly after conception (≤20 weeks) if prepregnancy was the only self-reported measure in GWG or shortly after delivery (<6 weeks) if women reported on total GWG (5,57,58,78,84). One study had a medium length of recall within 6 months of delivery (43), and one study had a long length of recall (4-12 years after delivery; (52)). Three short/medium recall studies used gold standard weight references (57,58,84). Correlation between self-reported and measured weight was low (r=0.6; (84)). Women tended to over-report their GWG by 1kg(58) to 3kg ((57); Table 3). One study assessed misclassification, finding that self-reported GWG overestimated population prevalence of excessive GWG (57).

In studies using alloyed gold standard weight references, misclassification ranged from 36%- 59% (5). Two studies assessed misclassification by both GWG adequacy and prepregnancy weight class (5,78). Both found fairly low misclassification for women with adequate GWG. One study additionally reported that misclassification for low GWG and high GWG varied by prepregnancy weight and race/ethnicity (5). Mean difference was smaller (0.33kg) in the one alloyed gold standard study reporting this measure (Table 3).

Findings from the study with a long recall were similar to those from studies with shorter lengths of recall, but misclassification for adequate and inadequate gainers was larger ((52);Table 3). Self-reported GWG also overestimated population prevalence of excessive GWG.

Ascertainment of self-reported GWG could either be reported directly by study participants or calculated from self-reported pre-pregnancy and delivery weight. Variation in error across these two ascertainment methods could be informative and are reported here. Two of the seven high-scoring studies used self-reported GWG based on the former method (52,78) and five on the latter (5,43,57,58,84). While the small number of direct-report studies precluded reliable assessment of variation across all measures of accuracy, correlation coefficients and misclassification did not vary greatly across this study characteristic (Table 3).

Postpartum Weight

We identified one study that investigated reporting error in self-reported postpartum weight, and it did not score highly on our validity assessment scale. More work is critically needed to appropriately assess reporting error in postpartum weight retention.

Variation Across Maternal Characteristics

A total of 18 studies reported variation of reporting error in pregnancy-related weight measures by different maternal characteristics (5,14,19,35,39,44,48,49,52,59,63,66,71,72,76,77,84). Studies did not consistently report the same maternal or demographic characteristics, so we focused on the few that were reported most consistently across studies. Because a small number of studies reported on variation by maternal characteristics, we included all studies regardless of validity assessment score.

Prepregnancy Weight

Studies most consistently reported on maternal age (14,19,39,59,63,66,71), socioeconomic status (SES) (14,39,59,63,66,72,84), and race/ethnicity

(5,14,19,35,39,59,63,66,72). In these studies, black women had higher reporting errors compared to white women (5,14,66,72), and reporting error did not vary by SES (14,59,63,66,72). Evidence on variation of reporting error by maternal age was equivocal. Three (14,39,66) of nine studies found that reporting error did vary by maternal age, but conflicted on which age groups had the greatest error.

Delivery Weight

The one study investigating variation in reporting error for delivery weight across maternal characteristics found variability across many characteristics (76). Reporting error was larger for women with higher prepregnancy and current BMI, non-adequate GWG (either low or high GWG), weight change from delivery to time of reporting, non-white race/ethnicity, unmarried marital status, lower education, unintended pregnancy, late or no prenatal visits, and subsequent pregnancy between measured and reported weight (76). The majority of subgroups with larger reporting errors were more likely to underreport delivery weight, but women with low GWG and women who gained weight between delivery and recall were more likely to over-report delivery weight (76).

Gestational Weight Gain

Evidence supporting variation of reporting error in GWG across maternal characteristics was mixed. The six studies we identified most consistently reported variation in reporting error by education (48,49,52), race/ethnicity (5,52,78), and parity (48,49). Error did not vary by maternal education (48,49), but did vary by race/ethnicity. Minority women were less likely to accurately report GWG (52,78). Two studies (48,49) investigated variation of reporting error by parity, and one found a higher correlation between self-reported and medically recorded GWG within first pregnancies (49).

Implications for Birth Outcomes

Qualitative Assessment

Table 4 displays results from our qualitative assessment of bias. Measures of accuracy were ranked based on whether they suggested a low, medium, or high risk of bias based on the criteria discussed previously in Table 2. Studies were limited to those that scored highly on our validity assessment scale. This qualitative assessment suggested that overall risk of bias from use of self-reported preprepregnancy weight was moderate to low, but varied by prepregnancy weight class (Table 4). Most studies (94%) reported correlation coefficients greater than 0.8 (14,19,35,38,39,48,53,58,60,63,66,68,72,84-86), and magnitudes of error less than 2.27kg (5lbs; 85.7%) suggesting a low risk of bias (5,14,19,35,38,39,43,48,53,56,58,63,68,72,82,84-86). However, variability in mean difference was greater than 2.27kg (5lbs) for 73.3% of studies (5,19,43,48,53,56,58,59,82,84,86), suggesting that while bias in associations between prepregnancy weight and outcomes may be small overall, it could be large for some sub-sets of the population.

Prevalence of misclassification indicated that risk of bias could be substantial for categorical measures of prepregnancy weight. Risk of bias was moderate to high in most of the 13 studies that reported misclassification (Table 4; (5,14,19,39,56,57,59,66,72,82,83,86)). Furthermore, misclassification varied by prepregnancy weight class. Sixty percent of studies on underweight women suggested a high risk of potential bias. Similarly, 66.7% of studies on overweight women suggested a high risk of bias. Studies for normal weight women all reported prevalence of misclassification that suggested moderate to low risk of bias. Among studies for obese women, only five studies found levels of misclassification suggesting a high risk of potential bias when using self-report.

For delivery weight, measures of accuracy suggested that risk of potential bias due to error from self-report was low for continuous measures (Table 4). All three studies reported mean differences that were less than 2.27kg (5lbs; (20,57,76)). However, risk of bias due to misclassification for categorical measures was high; both studies reported misclassification greater than 20% (20,76).

For GWG, measures of accuracy suggested that potential bias due to error in self-report was moderate to high. Mean difference was the only exception; 80% of studies reported mean differences less than 2.27kg (5lbs) suggesting low risk of bias (43,52,58,84). Nonetheless, studies reporting variability in error suggested that bias could be moderate to high (43,84), and studies reporting correlation coefficients suggested that bias could be high (52,84). Misclassification by relevant GWG adequacy groups also suggested that risk of bias could be high. Four studies reported misclassification prevalence greater than 20% when using self-reported measures (Table 4; (5,43,52,78)).

Empirical Assessment

Studies that empirically estimate bias in measures of association resulting from self-reported weight suggested that degree of bias was minimal. We identified eight studies (13,39,52,56,67,76,87,88) across seven cohorts that assessed magnitude of bias across a variety of pregnancy and birth outcomes. Among studies assessing bias in prepregnancy weight, three found that associations between prepregnancy weight and inadequate GWG, preterm birth, and small-for-gestational age were biased enough to change the interpretation of the association when self-reported weight was used (13,39,87).

One study assessed biases in associations between GWG and perinatal outcomes due to self-reported prepregnancy weight (67). They found that associations between inadequate GWG and both small-for-gestational age and gestational hypertension were biased for normal weight and obese women (Table 5). These authors also found that self-reported weight biased associations between excessive GWG and gestational hypertension for underweight, normal weight, and overweight women (Table 5). One study looking at self-reported delivery weight found that it biased associations between GWG and birth outcomes enough to change the interpretation of the associations for low GWG and low birth weight as well as high GWG and high birth weight (Table 5) (76).

For GWG, one study found that the association between excessive GWG and preterm birth was substantially changed due to bias from self-reported weight (52). Another study found that associations between adequate GWG and infant mortality were biased by self-reported weight measures (88) for underweight, overweight, and grade I obese women. Across all 20 associations that were biased by self-reported pregnancy-related weight, approximately half were biased away from the null, detecting an association not observed when appropriate reference weight measure were used (data available upon request).

While the 20 associations described above were substantially biased by use of self-reported pregnancy-related weight, they represent only 16.1% of the total associations evaluated for bias across identified studies. For the remaining associations (13,39,52,56,67,76,87,88), bias from use of self-reported weight measures was small and did not change associations between pregnancy-related weight and birth outcomes (Table 5).

DISCUSSION

Results of this systematic review confirmed that women underestimated their prepregnancy weight, and overestimated their GWG. Magnitude of error was relatively small, with high variability. Correlation was high between self-reported and objectively measured prepregnancy and delivery weight and moderate for GWG. However, misclassification was moderate to high for both prepregnancy BMI and GWG adequacy categories. We found relatively little evidence on reporting of delivery weight and not enough evidence for postpartum weight. We found evidence to suggest that minority groups and women with high BMI may be more likely to misreport all weight measures assessed, while evidence for maternal age, SES and maternal education was less consistent. Importantly, despite documenting these errors in self-reported maternal weight, a number of studies concluded that associations between self-reported prepregnancy weight or GWG and perinatal outcomes were similar when reporting error was considered in their analysis.

This is the first systematic review, to the best of our knowledge, to assess reporting error in multiple measures of weight over the course of pregnancy. Our findings agree with previous non-systematic reviews (9,17) which find that while correlation is high and mean difference is small (12), reporting error for self-reported weight varies across individuals within populations and across sociodemographic characteristics, such as age and education (17). Mean differences for self-reported pregravid weight range from -1kg to -2.4kg (18).   Our review adds new knowledge on misclassification of women by prepregnancy BMI weight classes or GWG adequacy categories. Misclassification led to differences in estimates of population prevalence of some pregnancy-related weight measures. This may be concerning for excessive GWG because self-reported measures overestimated prevalence, but more data is needed to confirm this observation. Such findings have implications for program planning and analyses using GWG as an outcome.

Reporting error in self-reported weight measures has been widely studied in non-pregnant women (4,32). Non-pregnant women underreport their weight by 0.1kg to 6.5kg with substantial variation in individual error (6,32). These averages are higher than we found in this review, perhaps because they are based on a wide range of ages, and self-reported weight error may increase with age (7,32,54,89). Our findings are more similar to those reported for non-pregnant women of childbearing age (-0.25kg (89) to -2.09 kg (6)). Correction techniques have been developed to address error in non-pregnant populations (10,32), and similarity in magnitude of reporting error for pregnancy-related weight suggests that these techniques may be applicable to pregnant women. However, future research should confirm that correction methods apply to weight measures at all key time points during pregnancy.

A main goal of investigating error in maternal pregnancy-related weight is to assess its role in biasing associations with health outcomes. Studies in our review that compared results using self-reported versus measured (39,52,56,67,76) or bias-adjusted measures (13,87,88) of maternal weight found that over 80% of associations between pregnancy-related weight and birth outcomes were not substantially biased. For example, inadequate GWG was associated with small-for-gestational-age births using both self-report and reference weight measures (52). However, studies have not assessed impacts of misclassification when pregnancy-related weight is the outcome, rather than the exposure. This is an important direction for future research because it impacts potential bias in associations between risk factors and maternal weight outcomes. Researchers should remain cautious when using self-reported GWG for population prevalence estimation and more work is needed to determine whether self-report ascertainment method (i.e. direct or calculated) influences magnitude of error and misclassification. We could not identify any studies that directly compared differences in validity for these two ascertainment methods and lacked enough data to make reliable comparisons. However, such information can inform correction methods for self-reported GWG going forward.

Overall, careful consider sample composition should be taken in order to anticipate true impact of bias in any particular study. Self-report remains an easy and efficient measurement approach for collecting weight information, and while it is reassuring that associations with health outcomes are usually unbiased despite reporting error, it is important to develop bias correction techniques for pregnancy-related weight measures. Especially, since clinical practice relies on accurate weight measures to appropriately allocate care, such bias correction methods can aid in correcting self-report that may occur within the clinical setting.

The current body of literature provides critical information on reporting error in self-reported pregnancy-related weight, but three major methodological limitations should be addressed to improve bias assessment going forward. First, many studies of prepregnancy weight rely on "alloyed gold standards" against which to compare women's self-report. One type is clinical records of maternal weight at first prenatal visit (e.g. (5,19,43)). This measure is almost always self-reported (1,5,19) and likely captures reliability of women's report rather than validity. Another type of alloyed gold standard for prepregnancy weight is measured weight at first prenatal visit. However, timing of first prenatal visit varies; in our review, it ranges from soon after conception to 22 weeks post-conception. Thus, this alloyed gold standard may be capturing women's actual weight gain since conception (19,63,90). We found some evidence that magnitude of error differed and misclassification was slightly larger when alloyed gold standards were used, indicating that future studies should more consistently use gold standard weight references. We recognize that ideal gold standard measures may be difficult to obtain for prepregnancy weight because women may not routinely engage with the medical system prior to pregnancy (90). Nonetheless, the development of preconception care programs (91,92) and use of electronic medical records that include prepregnancy weight (e.g. (39)) may aid with this. However, in the absence of these gold standard weight measures, researchers should give preference to alloyed gold standards based on measured weight at first prenatal visit. Adjustment algorithms based on rate of GWG and gestational age can then be applied to recover estimates of pre-pregnancy weight (e.g. (63)). Future work should focus on developing these adjustment algorithms for use when gold standard weight measures are not available.

Second, comprehensive assessment of reporting error in self-reported pregnancy-related weight is limited by incomplete reporting on multiple measures of accuracy. Many studies we identified reported only one or two measures of accuracy. Correlation coefficients were most commonly reported, but this measure may artificially represent agreement between the two measures, thus not capturing actual bias in reporting error. Furthermore, studies that included multiple measures of accuracy found that variability in error can be large and misclassification substantial even when correlation is high. Future studies should report multiple measures of accuracy, with particular focus on misclassification across clinically relevant prepregnancy weight and GWG adequacy categories, which are used to guide care and assess risk of poor birth outcomes in women.

Sampling design is the third methodological challenge. Many studies are not designed to assess reporting error, resulting in samples that are limited, post hoc, to women who happen to have both medically recorded and self-reported weight. This reliance on convenience samples may bias findings if reporting error varies across characteristics that are also associated with having complete medical records (93). Bodnar et al. (5) counteract this concern using a balanced design (30) to intentionally sample the population based on factors important for characterizing reporting error , including race/ethnicity and prepregnancy weight (5). More studies should follow this approach to ensure that appropriate sample sizes are available to assess validity of pregnancy-related weight across multiple maternal characteristics. Overall, studies addressing the methodological limitations described above will improve researchers' ability to assess bias due to self-reported pregnancy-related weight.

Our review is subject to a few limitations. First, we included only studies published in English. Second, despite our careful search strategy, it is always possible that we missed some relevant articles. Finally, we were unable to identify relevant studies on reporting error in postpartum weight retention. While this could reflect a gap in our search approach; it could also be due to lack of studies that distinguish postpartum weight from regular weight reported in non-pregnant women who happen to be post-delivery. Studies that use self-reported postpartum weight may assume that errors associated with it are similar to those in non-pregnant populations of women of reproductive age (e.g. (12,94)). However, such errors may differ due to postpartum lifestyle changes and stress associated with adding a new child to the family (95). Future work should address this gap.

CONCLUSION

While measuring weight using standardized measures is optimal and can be supported through the advent of electronic medical record data on weight before, during and after pregnancy (1), self-reported weight remains an easy and efficient option for maternal weight measurement. This systematic review investigated the accuracy and misclassification of self-reported, prepregnancy, gestational and delivery weight. While average error was small, misclassification still influenced population prevalence estimates of prepregnancy weight and GWG adequacy. However, reporting error did not seriously bias observed associations between pregnancy-related weight and birth outcomes. Nonetheless, in order to ensure that pregnancy-related weight risks can be more accurately quantified, it is imperative that bias correction techniques be developed based on more complete validity data. Future studies should be conducted in large, diverse, nationally representative samples of pregnant or recently pregnant women, using appropriate gold standards to assess multiple measures of accuracy and support this research need.

References:

1. Rasmussen KM, Yaktine AL, eds. Weight gain during pregnancy. Washington, DC: National Academies Press (US); 2009.

2. Margerison-Zilko CE, Rehkopf D, Abrams B. Association of maternal gestational weight gain with short- and long-term maternal and child health outcomes. *Am J Obstet Gynecol*. 2010;202(6):574.e1–574.e8.

3. Mamun AA, Mannan M, Doi SAR. Gestational weight gain in relation to offspring obesity over the life course: a systematic review and bias-adjusted meta-analysis. *Obesity Reviews*. 2013;15(4):338–347.

4. Willett W, Hu F. Anthropometric Measures and Body Composition. In: Willett W, ed. *Nutritional Epidemiology*. New York: Oxford University Press, USA; 2013:213–240.

5. Bodnar LM, Abrams B, Bertolet M, et al. Validity of Birth Certificate-Derived Maternal Weight Data. *Paediatric and Perinatal Epidemiology*. 2014;28(3):203–212.

6. Brunner Huber LR. Validity of Self-reported Height and Weight in Women of Reproductive Age. *Matern Child Health J*. 2006;11(2):137–144.

7. Engstrom JL, Paterson SA, Doherty A, et al. Accuracy of Self-Reported Height and Weight in Women: An Integrative Review of the Literature. *J Midwifery Womens Health*. 2003;48:338–345.

8. KUCZMARSKI MF, KUCZMARSKI RJ, NAJJAR M. Effects of age on validity of self-reported height, weight, and body mass index: Findings from the third National Health and Nutrition Examination Survey, 1988-1994. *J am Diet Assoc*. 2001;101:28–34.

9. Amorim AR, linne Y, Kac G, et al. Assessment of weight changes during and after pregnancy: practical approaches. *Maternal and Child Nutrition*. 2008;4:1–13.

10. Willett W. Correction for the Effects of Measurement Error. In: Willett W, ed. *Nutritional Epidemiology*. New York: Oxford University Press, USA; 2013:287–304.

11. Carter EB, Tuuli MG, Caughey AB, et al. Number of prenatal visits and pregnancy outcomes in low-risk women. *Journal of Perinatology*. 2015;36(3):178–181.

12. Gunderson EP. Childbearing and Obesity in Women: Weight Before, During, and After Pregnancy. *Obstetrics and Gynecology Clinics of North America*. 2009;36(2):317–332.

13. Bodnar LM, Siega-Riz AM, Simhan HN, et al. The impact of exposure

misclassification on associations between prepregnancy BMI and adverse pregnancy outcomes. *Obesity* [electronic article]. 2010;18(11):2184–2190. (http://doi.wiley.com/10.1038/oby.2010.25)

14. Park S, Sappenfield WM, Bish C, et al. Reliability and Validity of Birth Certificate Prepregnancy Weight and Height Among Women Enrolled in Prenatal WIC Program: Florida, 2005. *Matern Child Health J*. 2009;15(7):851–859.

15. Mendez DD, Doebler DA, Kim KH, et al. Neighborhood Socioeconomic Disadvantage and Gestational Weight Gain and Loss. *Matern Child Health J*. 2013;

16. Reichman NE, Hade EM. Validation of birth certificate data: A study of women in New Jersey's HealthStart program. *Ann Epidemiol*. 2001;11:186–193.

17. Harris HE, Ellison GT. Practical approaches for estimating prepregnant body weight. *J Nurse Midwifery*. 1998;43(2):97–101.

18. Cecchino NJ. Google Scholar. *J Med Libr Assoc*. 2010;98(4):320–321.

19. Lederman SA, Paxton A. Maternal Reporting of Prepregnancy Weight and Birth Outcome: Consistency and Completeness Compared with the Clinical Record. *Matern Child Health J*. 1998;2(3):123–126.

20. Alves E, Lunet N, Correia S, et al. Medical record review to recover missing data in a Portuguese birth cohort: agreement with self-reported data collected by questionnaire and inter-rater variability. *Gaceta Sanitaria*. 2011;25(3):211–219.

21. Armstrong BG. The effects of measurement errors on relative risk regression. *Am J Epidemiol*. 2005;132(6):1176–1184.

22. Spiegelman D. Approaches to Uncertainty in Exposure Assessment in Environmental Epidemiology. *Annu. Rev. Public. Health.* 2010;31(1):149–163.

23. Measuring agreement in method comparison studies. 1999;:1–26.

24. Selvin S. Variation and Bias. In: *Statistical Analysis of Epidemiological Data*. New York: Oxford Univ Press; 2004:40–74.

25. World Health Organization. Obesity: preventing and managing the global epidemic. Geneva: World Health Organization; 2000.

26. Hedderson MM, Gunderson EP, Ferrara A. Gestational Weight Gain and Risk of Gestational Diabetes Mellitus. *Obstet Gynecol*. 2010;115(3):597–604.

27. Willett W, Lenart E. Reproducibility and validity of food-frequency questionnaires. In: Willett W, ed. *Nutritional Epidemiology*. New York: Oxford University Press, USA; 2013:96–141.

28. Lash TL, Fox MP, Fink AK. Data sources for bias analysis. In: *Applying Quantitative Bias analysis to Epidemiologic Data*. New York: Springer-Verlang New York; 2009:33–42.

29. Spiegelman D. Validation Study. Chichester, UK: John Wiley & Sons, Ltd; 2005 7 p.

30. Holcroft CA, Speigelman D. Design of validation studies for estimating the odds ratio of exposure-disease relationships when exposure is misclassified. *Biometrics*. 1999;55:1193–1201.

31. Thomas D, Stram D, Dwyer J. Exposure measurement error: Influence of exposure-disease releationships and methods of correction. 1993;14:69–93.

32. Gorber SC, Tremblay M, Moher D, et al. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity Reviews*. 2007;8(4):307–326.

33. Hofler M. The effect of misclassification on the estimation of association: a review. *International Journal of Methods in Psychiatric Reserach*. 2005;14(2):92–101.

34. Herring SJ, Oken E, Haines J, et al. Misperceived pre-pregnancy body weight status predicts excessive gestational weight gain: findings from a US cohort study. *BMC Pregnancy Childbirth*. 2008;8(1):54.

35. Oken E, Taveras EM, Kleinman K, et al. Gestational weight gain and child adiposity at age 3 years. *Am J Obstet Gynecol*. 2007;196(4):322.e1–8.

36. Provenzano AM, Rifas-Shiman SL, Herring SJ, et al. Associations of Maternal Material Hardships During Childhood and Adulthood with Prepregnancy Weight, Gestational Weight Gain, and Postpartum Weight Retention. *Journal of Women's Health*. 2015;24(7):563–571.

37. Perng W, SM MWGM, DSc CSMM, et al. Annals of Epidemiology. *Annals of Epidemiology*. 2014;24(11):793–800.e1.

38. Ferrara A, Weiss NS, Hedderson MM, et al. Pregnancy plasma glucose levels exceeding the American Diabetes Association thresholds, but below the National Diabetes Data Group thresholds for gestational diabetes mellitus, are related to the risk of neonatal macrosomia, hypoglycaemia and hyperbilirubinaemia. *Diabetologia*. 2006;50(2):298–306.

39. Han E, Abrams B, Sridhar S, et al. Validity of self-reported pre-pregnancy weight and body mass index classification in an integrated health care delivery system. *Paediatr Perinat Epidemiol*. 2016;30:314–319.

40. Durmuş B, Arends LR, Ay L, et al. Parental anthropometrics, early growth and the risk of overweight in pre-school children: the Generation R Study. *Pediatric*

Obesity. 2012;8(5):339–350.

41. Gaillard R, Durmuş B, Hofman A, et al. Risk factors and outcomes of maternal obesity and excessive weight gain during pregnancy. *Obesity*. 2013;21(5):1046–1055.

42. Kirkegaard H, Stovring H, Rasmussen KM, et al. How do pregnancy-related weight changes and breastfeeding relate to maternal weight and BMI-adjusted waist circumference 7 y after delivery? Results from a path analysis. *American Journal of Clinical Nutrition*. 2014;99(2):312–319.

43. Nohr EA, Vaeth M, Baker JL, et al. Combined association of prepregnancy body mass index and gestational weight gain with the outcome of pregnancy. *Am J Public Health*. 2008;87:1750–1759.

44. Wright RJ, Fisher K, Chiu Y-HM, et al. Disrupted Prenatal Maternal Cortisol, Maternal Obesity, and Childhood Wheeze. Insights into Prenatal Programming. *Am J Respir Crit Care Med*. 2013;187(11):1186–1193.

45. Lakshmanan A, Chiu Y-HM, Coull BA, et al. Associations between prenatal traffic-related air pollution exposure and birth weight_ Modification by sex and maternal pre-pregnancy body mass index. *Environmental Research*. 2015;137(C):268–277.

46. Callaway LK, Mamun A, McIntyre HD, et al. Does a history of hypertensive disorders of pregnancy help predictfuture essential hypertension? Findings from a prospectivepregnancy cohort study. *Journal of Human Hypertension*. 2012;27(5):309–314.

47. Mamun Al A, Mannan M, O'Callaghan MJ, et al. Association between Gestational Weight Gain and Postpartum Diabetes: Evidence from a Community Based Large Cohort Study. *PLoS ONE*. 2013;8(12):e75679–9.

48. Tomeo CA, Rich-Edwards JW, Michels KB, et al. Reproducibility and Validity of Maternal Recall of Pregnancy-Related Events. *Epidemiology*. 1999;10(6):774–777.

49. Buka SL, Goldstein JM, Spartos E, et al. The retrospective measurement of prenatal and perinatal events: accuracy of maternal recall. *Schizophrenia Research*. 2004;71(2-3):417–426.

50. Puhkala J, Luoto R, Ahotupa M, et al. Postpartum Weight Retention is Associated with Elevated Ratio of Oxidized LDL Lipids to HDL-Cholesterol. *Lipids*. 2013;48(12):1227–1235.

51. Puhkala J, Kinnunen TI, Vasankari T, et al. Prevalence of Metabolic Syndrome One Year after Delivery in Finnish Women at Increased Risk for Gestational Diabetes Mellitus during Pregnancy. *Journal of Pregnancy*. 2013;2013(4):1–7.

52. McClure CK, Bodnar LM, Ness R, et al. Accuracy of Maternal Recall of Gestational Weight Gain 4 to 12 Years After Delivery. *Obesity*. 2011;19(5):1047–1063.

53. Catov JM, Abatemarco D, Althouse A, et al. Patterns of gestational weight gain related to fetal growth among women with overweight and obesity. *Obesity*. 2015;23(5):1071–1078.

54. Craig BM, Adams AK. Accuracy of Body Mass Index Categories Based on Self-Reported Height and Weight Among Women in the United States. *Matern Child Health J*. 2008;13(4):489–496.

55. Hickey CA, Cliver SP, Goldenberg RL, et al. Prenatal weight gain, term birth weight and fetal growth retardation among high-risk multiparous black and white women. *Obstet Gynecol*. 1993;81:529–535.

56. Holland E, Moore Simas TA, Doyle Curiale DK, et al. Self-reported Pre-pregnancy Weight Versus Weight Measured at First Prenatal Visit: Effects on Categorization of Pre-pregnancy Body Mass Index. *Matern Child Health J*. 2012;

57. Mandujano A, Presley L, Waters T, et al. Women's reported weight: Is there a discrepancy? *J Matern Fetal Neonatal Med*. 2012;25(8):1395–1398.

58. Sharma AJ, Vesco KK, Bulkley J, et al. Associations of Gestational Weight Gain with Preterm Birth among Underweight and Normal Weight Women. *Matern Child Health J*. 2015;19(9):2066–2073.

59. Masiero J, Stone JM, Simas TAM, et al. Accuracy of recalled compared with measured weight for the calcuation of prepregnancy body mass index [275]. *Obstet Gynecol*. 2015;125:88S–89S.

60. Vidal AC, Semenova V, Darrah T, et al. Maternal cadmium, iron and zinc levels, DNA methylation and birth weight. *BMC Pharmacology and Toxicology*. 2015;:1–9.

61. Mongraw-Chaffin ML, Anderson CAM, Clark JM, et al. Prepregnancy body mass index and cardiovascular disease mortality: The child health and development studies. *Obesity*. 2013;22(4):1149–1156.

62. Huang JY, Gavin AR, Richardson TS, et al. Are Early-Life Socioeconomic Conditions Directly Related to Birth Outcomes? Grandmaternal Education, Grandchild Birth Weight, and Associated Bias Analyses. *American Journal of Epidemiology*. 2015;182(7):568–578.

63. Yu SM, Nagey DA. Validity of Self-Reported Pregravid Weight. *Ann Epidemiol*. 1992;2(5):715–721.

64. Walker LO, Timmerman GM, Sterling BS, et al. Do low-income women attain their

pre-pregnant weight by the 6th week of postpartum? *Ethn Dis*. 2004;14:119–126.

65. Guendelman S, Pearl M, Kosa JL, et al. Association Between Preterm Delivery and Pre-pregnancy Body Mass (BMI), Exercise and Sleep During Pregnancy Among Working Women in Southern California. *Matern Child Health J*. 2012;17(4):723–731.

66. Shin D, Chung H, Weatherspoon L, et al. Validity of Prepregnancy Weight Status Estimated from Self-reported Height and Weight. *Matern Child Health J*. 2013;

67. Johnson J, Clifton R, Roberts J, et al. Pregnancy outcomes with weight gain above or below the 2009 institute of medicine guidelines. *Obstet Gynecol*. 2013;121(5):969–975.

68. Waage CW, Falk RS, Sommer C, et al. Ethnic differences in postpartum weight retention: a Norwegian cohort study. *BJOG: Int J Obstet Gy*. 2015;:n/a–n/a.

69. Lin X, Aris IM, Tint MT, et al. Ethnic Differences in Effects of Maternal Pre-Pregnancy and Pregnancy Adiposity on Offspring Size and Adiposity. *The Journal of Clinical Endocrinology & Metabolism*. 2015;100(10):3641–3650.

70. Scifres CM, Abebe KZ, Jones KA, et al. Gestational Diabetes Diagnostic Methods (GD2M) Pilot Randomized Trial. *Matern Child Health J*. 2015;:1–9.

71. Stevens-Simon C, McAnarney ER, Coulter MP. How Accurately Do Pregnant Adolescents Estimate Their Weight Prior to Pregnancy? *J Adolesc Health Care*. 1986;7(4):250–254.

72. Krakowiak P, Walker CK, Tancredi DJ, et al. Maternal Recall Versus Medical Records of Metabolic Conditions from the Prenatal Period: A Validation Study. *Matern Child Health J*. 2015;:1–11.

73. Badon SE, Dyer AR, Josefson JL, et al. Gestational weight gain and neonatal adiposity in the hyperglycemia and adverse pregnancy outcome study-North American region. *Obesity*. 2014;22(7):1731–1738.

74. Gould Rothberg BE, Magriples U, Kershaw TS, et al. Gestational weight gain and subsequent postpartum weight loss among young, low-income, ethnic minority women. *Am J Obstet Gynecol*. 2011;204:52.e1–52.e11.

75. Aubuchon-Endsley NL, Bublitz MH, Stroud LR. Biological Psychology. *Biological Psychology*. 2014;102:38–43.

76. Schieve LA, Perry GS, Cogswell ME, et al. Validity of Self-reported Pregnancy Delivery Weight: An Analysis of the 1988 National Maternal and Infant Health Survey. *Am J Epidemiol*. 1999;150(9):947–956.

77. Biro FM, Wiley-Kroner B, Whitsett D. Perceived and Measured Weight Changes

During Adolescent Pregnancy. *J Pediatr Adolesc Gynecol*. 1999;12:31–32.

78. Wright CS, Weiner M, Localio R, et al. Misreport of Gestational Weight Gain (GWG) in Birth Certificate Data. *Matern Child Health J*. 2012;16(1):197–202.

79. Rauh K, Gabriel E, Kerschbaum E, et al. Safety and efficacy of a lifestyle intervention for pregnant women to prevent excessive maternal weight gain: a cluster-randomized controlled trial. *BMC Pregnancy Childbirth*. 2013;13(1):1–1.

80. Macdonald-Wallis C, Tilling K, Fraser A, et al. Gestational weight gain as a risk factor for hypertensive disorders of pregnancy. *YMOB*. 2013;209(4):327.e1–327.e17.

81. Buescher PA, Taylor KP, Davis MH, et al. The Quality of the New Birth Certificate Data: A Validation Study in North Carolina. *Am J Public Health*. 1993;83:1163–1165.

82. Fattah C, Farah N, O'Toole F, et al. Body Mass Index (BMI) in women booking for antenatal care: Comparison between selfreported and digital measurements. *European Journal of Obstetrics & Gynecology and Reproductive Biology*. 2009;144(1):32–34.

83. Jeffs E, Sharp B, Gullam J, et al. Weight and height measurement: Potential impact in obstetric care. *Journal of the New Zealand Medical Associaiton*. 2014;127(1392):17–26.

84. Stevens-Simon C, Roghmann KJ, McAnarney ER. Relationship of self-reported prepregnant weight and weight gain during pregnancy to maternal body habitus and age. *J am Diet Assoc*. 1992;92:85–87.

85. He X, Hu C, Chen L, et al. The association between gestational weight gain and substantial weight retention 1-year postpartum. *Arch Gynecol Obstet*. 2014;290(3):493–499.

86. Niquini RP, Bittencourt SA, do Carmo Leal M. Conformity of pre-gestational weight measurement and agreement of anthropometric data reported by pregnant women and those recorded in prenatal cards, City of Rio de Janeiro, 2007–2008. *Rev Bras Epidemiol*. 2013;16(3):670–681.

87. Lash TL, Abrams B, Bodnar LM. Comparison of Bias Analysis Strategies Applied to a Large Data Set. *Epidemiology*. 2014;25(4):576–582.

88. Bodnar LM, Siminerio LL, Himes KP, et al. Maternal obesity and gestational weight gain are risk factors for infant death. *Obesity*. 2016;24(2):490–498.

89. Roth LW, Allshouse AA, Lesh J, et al. The correlation between self-reported and measured height, weight, and BMI in reproductive age women. *Maturitas*. 2013;76(2):185–188.

90. OHiggins AC, Doolan A, Mullaney L, et al. The relationship between gestational weight gain and fetal growth: time to take stock? *J Perinat Med*. 2014;42(4):409–415.

91. Paden MM, Avery DM. Preconception Counseling to Prevent the Complications of Obesity during Pregnancy. *American Journal of Clinical Medicine*. 2012;9(1):30–35.

92. Atrash HK, Johnson K, Adams MM, et al. Preconception care for improving perinatal outcomes: The time to act. *Matern Child Health J*. 2006;10:S3–S11.

93. Smith PC, Araya-Guerra R, Bublitz C, et al. Missing Clinical Information During Primary Care Visits. *JAMA*. 2005;293:565–571.

94. Gore SA, Brown DM, West DS. The role of postpartum weight retention in obesity among women: A review of the evidence. *Ann Behav Med*. 2003;26(2):149–159.

95. Phelan S. Pregnancy: a “teachable moment” for weight control and obesity prevention. *American Journal of Obstetrics and Gynecology*. 2010;202(2):135.e1–135.e8.